UNCLASSIFIED

AD 406155

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION ALEXANDRIA. VIRGINIA

CLASSIFICATION CHANGED
TO UNCLASSIFIED
FROM UNCLASSIFIED - LIMITED

TAB 64-21 DATE 1 Nov. 1964

UNCLASSIFIED

RADC-TDR-63-93 2 February 1963

AUTOMATIC ABSTRACTING

C107-3U1

TRW COMPUTER DIVISION
THOMPSON RAMO WOOLDRIDGE INC.
CANOGA PARK, CALIFORNIA

Contract No. AF 30(602)-2223
Engineering Change B

Prepared under the Sponsorship of the
Intelligence Processing Laboratory
Rome Air Development Center
Air Force Systems Command
United States Air Force
Griffiss Air Force Base
New York

ACKNOWLEDGMENT

The TRW team expresses appreciation to several members of the staff of Rome Air Development Center for invaluable assistance in the performance of this work. Mr. John McNamara of the Data Handling Section was Project Engineer on this study. Mr. Al Leavitt of the Human Engineering Laboratory assisted in the area of experimental design. Both of these men made particularly useful contributions to the project.

Key members of the TRW team were J. L. Kuhns, who conducted mathematical research, Dr. P. L. Garvin who assisted on dictionary compilation, L. Ertel, who performed the programming, and J. Brewer who performed content analysis.

H. P. Edmundson

H. P. Edmundson, Project Manager

ABSTRACT

~~This Final Report on~~ Automatic Abstracting presents a series of additions and refinements to the previous RADC contract study. During this contract ~~we have produced~~ two major results: An operating system and a research methodology. The operating system produces automatic abstracts via programs written for the IBM 7090 and the IBM 1401. The programming involved the preparation of an Edit Program which inputs the text of documents, a Cue Dictionary Program which inputs a fixed word list, and an Abstracting Program which selects and outputs sentences of the document. The research methodology proceeds from linguistic analysis of documents comprising a sample library, the compilation of dictionaries, the formulation of abstracting rules which are applied to new documents of an experimental library, and concludes with testing and evaluation of the final program and dictionaries on documents of a test library.

Examples are given of abstracts produced by the four basic methods. Cue method, Key method, Title method, and Location method. In addition, a combined method using Cue-Title-Location is exemplified as the preferred method. ~~Presented in a~~ separate volume are ~~211~~ examples of automatic ~~abstracts produced by this combined~~ method. Conclusions resulting from this study and recommendations for future research are presented.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF EXHIBITS

1. INTRODUCTION AND SUMMARY

1.1 STRUCTURE OF FINAL REPORT

The Final Report covering this contract consists of descriptions of the research performed, analysis of results obtained, copies of computer flow charts, together with operating instructions so that the automatic Abstracting Program can be use-tested. Program decks and dictionary decks have been submitted separately.

This final report on Automatic Abstracting systems consists of three levels of detail.

The first level presents in the broadest terms the statement of the problem and results obtained and is intended for the reader who is interested only in gross landmarks. Section 1 of this report treats this first level and is printed on colored paper.

The second level of detail covers the operating system and the research methodology in greater detail and is intended for the reader who has had some experience with the problem of automatic abstracting or allied problems such as automatic translation. Sections 2 and 3 of this final report cover this material.

The third and final level covers the same ground as levels 1 and 2, but in minute detail, and is intended to be read by programmers and computer specialists who need exact descriptions of formats and instructions in order to operate the abstracting system on a computer. The Appendix of this report consists of five sections which give these fine details.

In addition to the preceding three levels the final report is supported by 211 machine abstracts of documents taken from a corpus on exotic fuels.

1.2 THE PROBLEM

1.2.1 Goals

The purpose of the contract was to conduct an investigation and study to develop techniques for the automatic abstracting of textual information.

This present study has the following orientation:

(1) Continuation of the previous research by developing new and modifying previous computer routines for automatic abstracting.

(2) Experimentation and investigation of the effectiveness of the abstracting routines on new bodies of text.

(3) Development of external objective criteria to evaluate the different abstracting routines relative to each other.

For the purposes of this contract, abstract content was defined in terms of subject matter, purpose, methods, conclusions, generalizations, and recommendations; the assumption was made that as a minimum, such abstract content would, to a large degree, satisfy the screening function of abstracts. The research plan was to continue the previous research by developing new and modifying previous computer routines for automatic abstracting. These procedures were tested in a series of experimental cycles, each consisting of computer run, analysis of output and program correction.

The tests were conducted on a library of homogeneous subject matter and the various abstracting techniques were evaluated relative to each other by external objective criteria developed in the course of the contract. The end products are detailed working programs and ratings of the potential worth of the final methods and products.

1.2.2 Definition of an Abstract

In defining an abstract of a document, we must specify the following three aspects: content, form, and length. The problem of content in an automatic abstract is that of selecting or rejecting sentences of the original document so as to form an extract or abstract. The problem of form is that of deciding how these sentences so selected are finally presented to the reader in relation to the formatting of the title, authors, headings and subheadings, graphics, footnotes, and references. The problem of length is that of deciding how many words or sentences will constitute the final output according to fixed rules, variable rules, and thresholds of compactness.

It is currently believed that the notion of the abstract of a document is simple and generally understood; i. e., to every document there corresponds one abstract. Or to put it mathematically, the abstract A is a function of the document D, i. e., $A = f(D)$. Moreover, since A is really an extract, A is a subset of D, i. e., $A \subseteq D$.

However, on closer examination we see that a document can have many abstracts which differ from one another in their intended use. Thus the act of abstracting is definitely goal-oriented. With the realization that it is misleading to conceive of <u>the</u> abstract, it is proper now to speak of <u>an</u> abstract of a document. Thus an abstract is a function of two quantities, the document D and the use U, i. e., $A = f(D, U)$.

Despite the fact that the preceding observation is simple and intuitively acceptable, its consequences are neither of these. In fact, it provides the foundation for the proper solution to the problem of automatic abstracting. Because of the number of alternative uses, it is necessary to define abstract content explicitly in terms that are use-oriented. This definition must be expressed by machine criteria. This requires detailed specification far beyond what intuitively might have been expected. Thus, we eliminate arguments over what is an abstract by replacing useless generalities with specific operational criteria.

This problem is closely related to the section of this report devoted to evaluation. It involves questions of the existence of a completely general definition of an abstract versus that of many specific definitions.

This leads us to the concept of a tailor-made abstract, in the sense that an individual will be able in future automatic systems to specify more accurately what he wants in an abstract. Moreover, this feature distinguishes automatic abstracting from automatic translation. It is generally believed that, aside from minor stylistic variations, there is only one translation of a document. On the other hand, we have seen that a document can have many different abstracts. This difference is fundamental to the problem of evaluating the quality of automatic abstracts.

1.2.3 Specific Tasks

The general problem of defining an abstract must next be translated into a set of specific tasks that will lead to an operating system. In light of the previous research that produced automatic abstracts from documents in a heterogeneous corpus, the shift to a new computer system and to a new corpus necessitated carrying out the following specific tasks:

(1) Convert previously developed computer routines to facilitate the abstracting experiments proposed for the new effort.

(2) Improve and modify existing abstracting routines; develop new linguistic factors for incorporation in abstracting routines; design abstracting experiments.

(3) Develop evaluation procedure.

(4) Select corpus for the new Experimental and Test Libraries, reproduce, pre-edit, and keypunch this corpus.

(5) Incorporate improved abstracting routines in computer program.

(6) Manually prepare extracts of documents from the Experimental Library; program mechanizable features of the evaluation procedures.

(7) Perform abstracting experiments on Experimental Library; evaluate and modify program.

(8) Perform abstracting experiments on the Test Library and evaluate abstracts produced.

(9) Prepare program flow charts and operating instructions.

1.3 PROJECT RESULTS

In accordance with the goals of the automatic abstracting project, this final report presents two aspects, an operating system and research methodology. The former is intended to provide a computation center with sufficient information to initiate an operating system for purposes of auotmatic abstracting, while the latter presents the research methodology initiated and developed during the contract, whose final result is the operating system.

1.3.1 Operating System

The operating system is presented first. The cardinal feature of the operating system is its flexibility. By this, we mean that the system has been parameterized whenever possible in order to allow easy modification by other users and for other purposes. This has been accomplished by making it possible to modify both externally and internally stored word lists, to re-adjust weights assigned to words, to permit the use of 15 combinations of the four basic methods of abstracting, and to alter the truncation thresholds which determine the length of an automatic abstract.

The operating system comprises:

(1) Two Corpora

Heterogeneous Corpus: 200 documents comprising several different subject fields (Physical Science, Life Science, Humanities, and Information Science).

Exotic Fuel Corpus: 200 ASTIA documents concerning the chemistry of exotic fuels.

Both corpora exist on punched cards and magnetic tape.

(2) <u>Four Computer Programs for Automatic Abstracting</u>

Cue Routine: weights sentences according to match of text words with Cue Dictionary.

Key Routine: weights sentences according to frequency of word occurrence.

Location Routine: weights sentences according to their location in document and matches text words with Heading Dictionary.

Title Routine: weights sentences according to match of text words with title and subheading words.

These programs were designed with experimentation in mind.

(3) <u>Computer Routines to Facilitate Experimentation</u>

Concordance Program: a program to generate a concordance of a document (with context displayed in the output) and classify words and sentences according to the number of times they were used in manual extracts (see Reference 7).

Cue Dictionary Program. a program to feed the Cue Dictionary into the Abstracting Program.

1.3.2 <u>Research Methodology</u>

This Final Report stresses the research performed during the second phase of the automatic abstracting contract. It will be recalled that the corpus used in the first phase of the contract was heterogeneous in that its documents came from the fields of political science, sociology, political affairs, astronomy, physics, etc. On the other hand, the corpus examined during the second phase of the contract was strictly homogeneous in that all the documents therein concerned the chemistry of exotic fuels.

This switch to the Exotic Fuel Corpus necessitated some changes in the words in the basic Cue Dictionary. Moreover, since these exotic fuel documents were contract reports obtained from ASTIA they were on the average more uniform with regard to size, format, and organization than documents of the earlier heterogeneous corpus. The factors of uniform length, shorter length, standard form, and technical writing style might cause the curious reader to wonder if the system originally based on a heterogeneous corpus would also serve for this homogeneous one and secondly, whether the set of changes based on the Exotic Fuel Corpus would make it impossible for the final operating system to handle properly the heterogeneous corpus. The answer to these questions cannot be completely given at this time; however, experimentation will settle the matter. It should also be mentioned that even while relying on the factors noted above, constant attention was paid toward producing an operating system which would work equally well for both the Exotic Fuel Corpus and any heterogeneous corpus. Determination of the extent to which this is true awaits further experimentation.

The research methodology comprises:

(1) A study of the extracting behavior of humans.

(2) A general formulation of the abstracting problem and its relation to the problem of evaluation. Four methods of evaluation that have been studied are statistical correlations, information content tests, retrievability tests, and judges' ratings (see Reference 7).

(3) A mathematical and logical study of the problem of assigning ranking numbers to sentences.

(4) A set of abstracting experiments based on cyclic improvement.

1.4 CONCLUSIONS AND RECOMMENDATIONS

1.4.1 Conclusions

The conclusions of this final report will be listed under two headings: operating system, and research methodology.

Operating System

(1) An operating system has been developed which produces automatic abstracts on the IBM 7090 computer by four distinct methods.

(2) The system abstracts technical documents whose lengths do not exceed approximately 3000 words.

(3) The computer abstracts at the rate of at least 7800 words per minute on a corpus of 29, 500 words (see Exhibit 1).

(4) The total system cost (edit, abstracting, system output) is approximately 1. 5 cents per word of which keypunching costs 1. 0 cents per word (see Exhibit 1).

(5) The operating system produces abstracts of sufficiently high quality to satisfy the screening function.

Research Methodology

(1) A definition of an abstract has been developed which leads to a more uniform target abstract prepared by humans and permits the creation of automatic abstracts based on machine-recognizable clues.

(2) The techniques for pre-editing and keypunching a new corpus have been rendered routine.

(3) Investigation has resulted in a research methodology which can be applied in the development of additional dictionaries, weighting systems, and evaluation techniques.

(4) The principle of attaining flexibility through parameterization has been verified.

(5) The method of research, by means of experimental cycles, has been shown to be effective and reliable.

(6) A method of evaluation by judges ratings of similarity has been developed. The resulting evaluation shows that the machine abstracts have a 66% degree of similarity with the target abstracts.

1. 4. 2 Recommendations

The recommendations of this final report will also be listed under two headings.

Operating System

(1) That the operating system be modified so as to abstract documents up to approximately 10, 000 words in length.

INPUT

Operation	Time (min/word)	Cost (dollar/word)
Pre-edit	.01	.0003
Keypunch[1]	.03	.01

ABSTRACTING METHODS[2]

Method	7090 Time[3]	Speed	
	min	words/min	min/word
L	1.3	23,000	.00004
T	1.5	20,000	.00005
C	2.3	13,000	.00008
K	3.0	9,800	.00010
L-T-C	3.1	9,500	.00011
L-T-C-K	3.8	7,800	.00013

PROCESSING AND OUTPUT

Operation	7090		1401		Total	
	Time[3] (min/word)	Cost[4] (dollar/word)	Time (min/word)	Cost[5] (dollar/word	Time (min/word)	Cost (dollar/word)
Edit [6]	.00013	.0011	.0023	.0019	.0024	.0030
Abstracting[6] (Method C-T-L) Without Vertical Listing	.00011	.00091	.00046	.00038	.00057	.0013
With Vertical Listing	.00020	.0017	.0028	.0023	.0030	.0040

(1) Includes verifying, sequencing, and interpreting.
(2) Estimates based on a sample of 29,500 words in 20 documents of Batch B. Add 2.8 minutes for vertical listing.
(3) Estimates do not include operator set-up time and program read-in-time. Add 0.4 minutes for read-in-time.
(4) 7090 time costs $8.30 per minute.
(5) 1401 time costs $0.833 per minute.
(6) Estimates based on a sample of 29,500 words.

Exhibit 1. Time and Cost Estimates

(2) That the operating system be reprogrammed as a more efficient program in order to reduce operating time and costs.

(3) That the method of inputting text be improved as to cost and speed.

(4) That the operating system be augmented to handle abstracting clues found in phrases, clauses, captions, footnotes, and references.

Research Methodology

(1) That any future research adhere to the principle of parameterization in order to gain flexibility.

(2) That the problem of capturing chemical and mathematical symbols in machine form be investigated more thoroughly.

(3) That investigations be made so as to reduce both the number and size of experimental cycles via more powerful statistical techniques.

(4) That further research and experimentation be conducted on evaluation methods in order to increase their speed, simplicity, and discrimination.

1.4.3 Outlook

In spite of the problems highlighted earlier it is felt that automatic abstracts can be defined, programmed, and produced in an operational system so as to compete with present human abstracting. That the future automatic abstracts may be different both in content and appearance from classical ones seems clear. However, it is not thought that users will be materially inconvenienced in having to adapt to a new format. Further research needs to be performed in this area of linguistic data processing, but the general outlines of the goal are being seen clearly for the first time.

2. OPERATING SYSTEM

By system aspects we refer to the functional specification of each of the steps in the automatic abstracting system. In general these steps are: (1) editing of the textual input, (2) selecting abstracting method, and (3) inputting of dictionary. Further details are shown in Exhibit 2.

The operating system consists of three main programs:

(1) The Edit Program consists of 1707 instructions and occupies 30K of 7090 core using STL's system B.

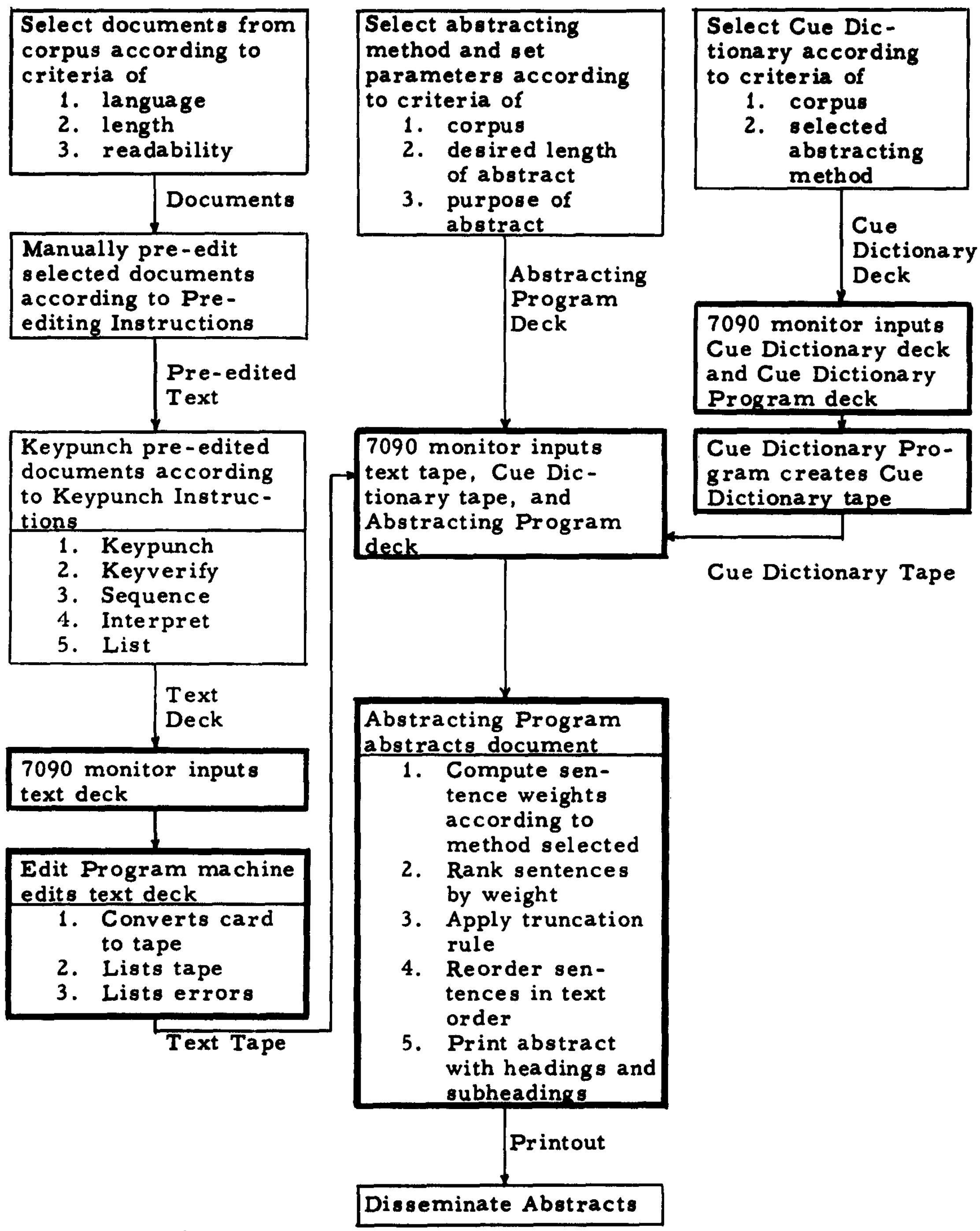
(2) The Cue Dictionary Program tape consists of 704 instructions and uses 23K of 7090 core.

(3) The Abstracting Program has 2427 instructions and uses 32K of 7090 core.

2.1 INPUT

2.1.1 Pre-editing

Certain subject fields introduce difficulties in the pre-editing step, since at the present time we must keypunch the original document. Moreover, even if a print reader were available, it would probably not be able to read a sufficient number of fonts for many years to come. This means that the text must be manually pre-edited according to a set of Pre-edit Instructions (see Appendix, Section 1). The creation of these instructions is not trivial because it is precisely at this step that we may choose to retain or ignore critical content and format clues which, once lost, can never be restored by any programming tricks. The Pre-edit Instructions must cover problems of formatting, graphics, special symbols, special alphabets. As an example, Exhibit 3 presents a reproduction of the pre-edited version of Document 6 from the Exotic Fuel Corpus. The notations entered are defined and discussed in the Appendix, Section 1.



Note: Heavy boxes denote machine operations.

Exhibit 2. Operating System

~~PHILLIPS PETROLEUM COMPANY - [illegible] DIVISION - REPORT [illegible]~~

~~SPECIAL REPORT~~

~~NAVY CONTRACT NOas [illegible], Amendment 7~~

∆ EVALUATION OF THE EFFECT OF DIMETHYLAMINE BORINE AND SEVERAL OTHER ADDITIVES ON COMBUSTION STABILITY CHARACTERISTICS OF VARIOUS HYDROCARBON TYPE FUELS IN PHILLIPS MICROBURNER ∆/AD57730/∆

R. L. Brace

∆∆∆∆ — SUMMARY

At the request of the Navy Bureau of Aeronautics, Phillips Petroleum Company undertook the evaluation of dimethylamine borine as an additive for improving the combustion characteristics of aviation gas turbine type fuels. Because of the small amount (100 grams) of dimethylamine borine received from Callery Chemical Company, this evaluation has been limited to the measurement of its effect on the flash-back characteristics of three pure hydrocarbons (toluene, normal heptane and benzene) in the Phillips Microburner. Dimethylamine borine concentrations of from 0.1 to 1.0 per cent by weight were evaluated.

For comparative purposes two common ignition additives (amyl nitrate and cumene hydroperoxide) were also evaluated during this study, as well as concentrations up to 20 per cent by weight of propylene oxide - a relatively high flame velocity fuel. Previous studies in Phillips 2 inch turbojet engine type combustor had indicated that such materials could substantially increase the maximum rate of heat release attainable, especially with low performance fuels such as the iso paraffin type hydrocarbons - particularly when operating under severe conditions for combustion (i.e., high air flow velocity or low combustion pressure).

The assumption has been made in this fuel evaluation that the greater the allowable heat input rate for a given velocity, the greater the degree of combustion stability. On this basis, the data indicate that all the additive materials tested caused an increase in stability performance; a fuel of relatively low performance such as toluene being benefited to a greater extent than a high performance fuel such as normal heptane. These data are in agreement with previous additive studies by Phillips. With respect to the dimethylamine borine, its effect as a fuel additive was noteworthy; 0.1 weight per cent in toluene being equivalent to 20 per cent by weight of added propylene oxide. In general, additive concentrations of one per cent by weight in the several pure hydrocarbons which normally differed quite widely in performance, produced uniformly superior combustion stability characteristics as measured using the Phillips Microburner.

∆∆∆∆ — I. INTRODUCTION

At the request of the Navy Bureau of Aeronautics the Jet Fuels Group has evaluated the effects of the addition of small amounts of dimethylamine borine ($(CH_3)_2NH{:}BH_3$) on the combustion stability performance of several hydrocarbon fuels. The dimethylamine borine was supplied to Phillips by the Callery Chemical Company.

Due to the small quantity of this material obtained the evaluation was conducted in the Phillips Microburner (Model 1A) which is a slightly modified version of the original Phillips Microburner (Model 1).

∆∆∆∆ — II.A DESCRIPTION OF PHILLIPS MICROBURNER (MODEL 1A)

The design of the original Phillips Microburner (Model 1) was discussed in detail in Reference 1. The only significant change in burner detail was the substitution of a #0 LeRoy Lettering Pen for the #3 LeRoy pen fuel nozzle in the Model 1. The details of the Model 1A are shown in Figure 1. #50 #53

∆∆∆∆ — III. DESCRIPTION OF TEST APPARATUS

The fuel system on the Phillips Microburner (Model 1) was designed so as to handle small quantities of highly corrosive substances. In the present evaluation it was not necessary to consider the effect of corrosion, consequently a continuous flow system providing greater flexibility and easier handling was incorporated which requires only slightly more fuel per test than the original. This new system includes a source of nitrogen gas, a surge tank, a pressurized fuel tank, a small capacity Brooks Rotameter and a flow control valve.

The air heater circuit was modified so as to include a mercury relay and bi-metallic thermoswitch thus providing automatic air temperature control.

The details of these modifications and of the test apparatus are shown in schematic in Figure 2.

∆∆∆∆ — IV. DESCRIPTION OF TEST FUELS

The chemical and physical properties of the hydrocarbon fuels tested are summarized in Table 1. These fuels represent variations in chemical structure which will in turn provide indices of both good and poor combustion stability performance.

The additives evaluated, dimethylamine borine, propylene oxide, amyl nitrate, cumene hydroperoxide, were blended into the hydrocarbon fuel by weight in concentrations ranging from 0.1 to 20 per cent.

∆∆∆∆ — V. TEST PROCEDURE

The conditions at which this study was made in the Microburner (model 1A) were: #T.

Air Pressure	20 psig
Inlet Air Temperature	300 ± 20 F
Fuel Pressure	30 psig
Fuel Flow x 10	3.85 to 6.82 pps
Air Flow x 10	~~3.77 to 6.69 pps~~

Once the air temperature was established the air flow was set at the desired rate and the fuel turned on. Ignition of the then fuel-rich mixture was accomplished by applying a lighted, portable propane torch to the top of the burner tube. The fuel flow was then gradually decreased until the flame holding at the burner tube rim flashed back into the tube. At the point of flashback the following were recorded:

- Cold Air Temperature
- Inlet Air Temperature
- Air Rotometer Reading
- Fuel Rotometer Reading

After checking the point at least once more the air flow was increased another increment and the process repeated.

Exhibit 3. Reproduction of Pre-edited Version of Document 6.

VI. RESULTS

The results of the evaluation of flashback limits of the fuels and fuel-additive blends are summarized in Table 2 and shown graphically in Figures 3 through 8. The recorded data were converted to values of heat input rate and a reference velocity at flashback - thus the flashback limits (or combustion stability characteristics) are established on the basis of an allowable heat input rate. The heat input rate is determined by the heating value of the fuel per pound of air or

HIR - lower heating value x $\left(\frac{\text{fuel flow}}{\text{air flow}}\right)$/lb Δ/FUEL FLOW/AIR FLOW/Δ

The reference velocity is determined by the air flow conditions at entry to the burner tube - neglecting the mass of the fuel particles.

The region of stable operation is defined as the state of flashback - the conditions of combustion where the flame would become anchored to a flame holder - as in stable gas turbine or ram jet combustor operation - if the flame holder were provided in the burner tube.

VII. DISCUSSION

The evaluation of the effects of the additives under consideration on the combustion stability characteristics of the fuel blends tested were interpreted from a plot of heat input rate at flashback against reference velocity. The assumption is made that the greater the allowable heat input rate at a given velocity, the greater the degree of stability.

Prior to testing the additive blends a base line of fuel performance was established by testing four relatively pure hydrocarbon fuels - each of different chemical structure but of comparable volatility. The relative performance of these fuels is shown in Figure 3, which agrees with the relative performance obtained in the original Microburner study. (ref. 1)

Propylene oxide was previously shown to provide an increase in combustion stability performance (ref. 1), consequently, various blends of toluene and propylene oxide were prepared in concentrations of 1, 5, 10, and 20 per cent by weight. The effect of the addition of propylene oxide to toluene on the flashback limits are shown in Figure 4. The addition of propylene oxide to toluene shows significant increases in allowable heat input rate. This evaluation served primarily as a reference plane of performance improvement.

Previous work conducted in the Phillips 2 inch combustor (ref. 2) indicated that some additives caused a significant increase in the performance of a low rating fuel while these same additives did not substantially effect the higher rating fuels. Therefore, in addition to the evaluation of the fuel blends containing 0.1, 0.5, and 1 per cent dimethylamine borine in toluene (shown in Figure 5) the same additive was tested in n-heptane (shown in Figure 6). Verification of the prior results was further established by testing two of the additives previously evaluated (ref. 2). The effects of these additives blended with n-heptane and toluene are shown in Figures 6 and 7 respectively.

To more definitely establish the effect of dimethylamine borine on the flashback limits of aromatic type hydrocarbon fuels, benzene with 0.84 weight per cent additive was evaluated as shown in Figure 8. The increase in allowable heat input rate with benzene-plus-additive is insignificant. The "boost" effected with additive in toluene is about four times that of the increase afforded with benzene, indicating the phenomena to be one of general performance level rather than a characteristic of aromatic type hydrocarbons.

In order to compare the additives evaluated, the allowable heat input rates of the fuel-additive blends at a constant velocity of 12 fps (arbitrarily selected) were plotted against additive concentration - shown in Figure 9. Examination of these curves shows dimethylamine borine to provide the greatest increase in stability performance and propylene oxide the least. All four additives indicated their addition to be subject to the effect of diminishing results upon further addition - that is, their effect was not essentially a blending effect.

Mention should be made of the fact that during the combustion of the dimethylamine borine-hydrocarbon fuel blends no noticeable odors or smoke were observed. The only visual difference encountered was the presence of a bright green flame instead of the usual blue hydrocarbon flame.

VIII. CONCLUSIONS

1. The addition of dimethylamine borine in concentrations of one per cent by weight to jet fuel type hydrocarbons resulted in a uniformly high level of combustion stability performance as measured by Phillips Microburner.

2. The addition of relatively large amounts of propylene oxide to toluene were necessary to provide significant improvement in stability performance as indicated by increases in allowable heat input rates.

3. The addition of additive concentrations (up to 1 per cent) of amyl nitrate, cumene hydroperoxide, and dimethylamine borine all resulted in improved stability performance; the greatest increases were shown when blended with a fuel of poor performance characteristics - such as toluene. Beneficial effects were appreciably less when blended with a fuel of good performance characteristics - such as in n-heptane.

IX. RECOMMENDATIONS

Based on the evaluation of the effects of additives on the flashback limits of the additive-fuel blends tested in the Microburner (Model 1A) it is recommended that dimethylamine borine should be further investigated. This future work should include study of combustion stability and combustion efficiency effects in the Phillips 2 inch combustor and an investigation of its influence on combustion cleanliness.

DOC'T ENDS HERE

X. REFERENCES

1. Brace, R. L., "Evaluation of the Performance Characteristics of the Phillips Microburner," Phillips Petroleum Company Report 403-23-53R, 1953.

2. Britter, S. C., Schirmer, R. M., and Fox, H. M., Final Report - Navy Contract NO a(s) 9596 "A Design Study for Equipment to Evaluate Performance of Aircraft Gas Turbine Fuels." Phillips Petroleum Company Report 763-49R, 1949.

Exhibit 3. Reproduction of Pre-edited Version of Document 6 (Continued)

2.1.2 Keypunching

Despite the fact that keypunch operators quickly adapt to new challenges it is necessary to prepare a set of Keypunch Instructions (see Appendix, Section 2). These instructions are based upon the Pre-edit Instructions and are subject to the boundary conditions imposed by present input and output hardware. They contain rules of sufficient generality to cover a wide variety of textual situations and should also be supported by appropriate examples. The purpose of these Keypunch Instructions is to relieve, to the maximum extent possible, the keypunch operator of making decisions.

The manually pre-edited text is put into machine-readable form by the following sequence of operations:

(1) Check pre-editing by scanning of document.

(2) Keypunch one line of text to a card on IBM 024 at 100 cards per hour.

(3) Keyverify on IBM 056.

(4) Sequence cards.

(5) Interpret cards.

(6) List in 2 parts.

(7) Spot check original document against machine listing.

2.1.3 Edit Program

The Edit Program creates the text tape which is used by the Abstracting Program.

Text written in free format is punched on cards according to the Keypunch Instructions (see Appendix, Section 2). The Edit Program interprets these cards, recognizing title, heading, and author cards; paragraphs, sentences, number of words in a sentence, punctuation, capitalization, etc. Each text word takes up five computer words (see Appendix, Section 3). A total of 1023 text words make up a record on the BCD output tape. The first word of each record contains the binary count of text words in a record. The last five words of each record contain zero, therefore, each record of the output tape contains 5121 BCD words.

The routine also recognizes input errors (see Output Error List, Appendix, Section 5). The errors are printed on the system output tape along with a vertical listing of text.

2.1.4 Cue Dictionary Program

The Cue Dictionary Program creates a BCD dictionary tape to be used by the Abstracting Program when it is determining Cue and Key words.

The input to the routine consists of at most 1000 words and their weights; Bonus words have positive integer weights ≤ 99, Null words have zero weights, and Stigma words have negative integer weights ≥ -99. The words are punched one per card (see Input Format, Appendix, Section 3). The words must be in alphabetical order.

The output consists of a BCD tape which contains only one record. Each entry takes four computer words (described in Output Format page, Appendix, Section 3). Certain input errors can be detected, and a notice will be printed on the system output tape. Errors are described in the Appendix, Section 5. A BCD listing of the output tape is also output on the system output tape.

2.2 PROCESSING

Programs

It is useful to separate the total system into three major operating programs: Edit Program, Cue Dictionary Program, and Abstracting Program. In addition to these operating programs various research programs were written during the research phase.

Based upon the theoretical model or structure underlying the abstracting system, decisions were made how best to use a mixture of computing routines and table-lookup routines. The abstracting system provides for the modification of the various parameters that are incorporated in the programming steps or that are stored in the tables. This allows discoveries made during research to be transformed easily into improvements in the computer program.

Tables

The success of an automatic abstracting system depends materially upon two different aspects. The first is the general system of abstracting as given by the sequence of programming operations. An example is the Cue method. The second aspect is the specific entries in stored tables. An example of a stored table is the Cue Dictionary of 1000 words that act either as Bonus words which signal the importance of a sentence or as Stigma words that signal the nonsignificance of a sentence. Such a table includes, in addition to the word, a code indicating its semantic function and its importance weight. Another kind of table is set aside to retain the title, author, and section headings. The programmer is presented with the considerable problem of juggling sections of the core memory so as to accommodate the input text, the program, and the tables.

2.2.1 Cue Method

The Cue method has as its source of machine recognizable clues the general characteristics of the corpus that are provided by the body of the documents (see Section 3, Rationale of the Four Basic Methods). The Cue method is based on a Cue Dictionary of certain function words apt to appear in the body of the document (see Section 3, Rationale of the Four Word Lists). There are three classes of Cue words in the Cue Dictionary: Bonus words, Stigma words, Null words. Thus, the Cue Dictionary can be conceived as consisting of three dictionaries: the Bonus Dictionary, Stigma Dictionary, and Null Dictionary. Bonus words are defined as those words of the Cue Dictionary that indicate that the sentence in which they appear should be in the abstract; therefore, Bonus words are assigned positive weights. Stigma words are defined as those words of the Cue Dictionary that indicate that the sentence in which they appear should not be in the abstract; therefore, Stigma words are assigned negative weights. Null words are defined as those words of the Cue Dictionary which are irrelevant to selection of the sentence for the abstract; therefore, Null words are assigned zero weights. The Null Dictionary was created and maintained for two reasons: (1) it constitutes a dictionary of common words that may upon further research be transferred to the Bonus or Stigma Dictionaries for the Cue method; (2) words that do not match this dictionary are made candidates for special attention in the Key, Title, and Location methods.

To illustrate the content (words and weights) and format of the Cue Dictionary a portion of it is presented in Exhibit 4.

The computation of sentence weights by the Cue method is detailed in the following steps which also show the present choice of parameter values:

(1) Compare each word of text with the Cue Dictionary.

(2) Tag all Bonus words with the weight b_i (= + 10).
Tag all Stigma words with the weight s_i (= - 10).
Tag all Null words with the weight n_i (= 0).

(3) Compute the Cue weight C of each sentence by summing its Cue-word weights b_i, s_i, and n_i.

(4) Rank all sentences in decreasing weight order.

(5) Select all sentences whose rank order is less than the threshold S(= 25 percent of the total number of sentences).

(6) Select all headings.

START	CC-PBC	WEIGHT	COUNT	AC1	AC2	WORD
77	000012	000010	000004			EXPEDIENT
77	000012	000010	000004			EXPERIMENT
77	000012	000010	000004			EXPERIMENTAL
77	000012	000010	000004			EXPERIMENTATION
77	000012	000010	000004			EXPERIMENTS
77	000012	000010	000004			EXPLAIN
77	000012	000010	000004			EXPLAINED
77	000012	000010	000004			EXPLAINING
77	000012	000010	000004			EXPLAINS
77	000012	000010	000004			EXPLANATION
77	000012	-00010	000004			EXPLANATIONS
77	000012	000010	000004			EXPLANATORY
77	000012	000010	000004			EXPLORE
77	000012	000010	000004			EXTENSIVE
77	000012	000010	000004			EXTRAORDINARY
77	000012	000010	000004			EXTREME
77	000012	000010	000004			EXTREMELY
77	000012	000010	000004			FAILED
77	000012	000010	000004			FAILURE
77	000012	000010	000004			FAINTER
77	000012	-00010	000004			FAIRLY
77	000012	000010	000004			FARTHER
77	000012	000010	000004			FASTER
77	000012	000010	000004			FAVOR
77	000012	000010	000004			FAVORABLE
77	000012	000010	000004			FEASIBLE
77	000012	000010	000004			FEW
77	000012	-00010	000004			FIG
77	000012	-00010	000004			FIGURE
77	000012	-00010	000004			FIGURES
77	000012	000010	000004			FIND
77	000012	000010	000004			FINDING
77	000012	000010	000004			FINDINGS
77	000012	000010	000004			FIRST
77	000012	000000	000004			FIVE
77	000012	000000	000004			FOR
77	000012	-00010	000004			FORMERLY
77	000012	000010	000004			FORMULATION

Exhibit 4. Section of Cue Dictionary

(7) Merge the selected sentences under their proper headings.

(8) Output title, authors and the results of step (7).

2.2.2 Key Method

The Key method has as its source of machine-recognizable clues the specific characteristics of the body of the document (see Section 3, Rationale of the Four Basic Methods). The Key method is based on a Key Glossary of content words taken from the body of the document (see Section 3, Rationale of the Four Word Lists). Key words which appear in the Key Glossary are created for each document to be abstracted by making statistical counts of the frequency of occurrence of all words in the document except those that are found in the fixed Cue Dictionary. The words resulting are ranked in decreasing frequency order and all those appearing above a pre-established threshold are defined as Key words.

The computation of sentence weights by the Key method is detailed in the following steps which also show the present choice of parameter values:

(1) Compare each word of text with the Cue Dictionary.

(2) Create table of distinct nonmatching text words (i. e., Key-word candidates).

(3) Compute frequency k_i of each Key-word candidate.

(4) Sort table in decreasing frequency order.

(5) Create Key Glossary from all Key-word candidates whose total frequency (when summed in decreasing frequency order) exceeds the threshold W(= 10 percent of the total number of word occurrences in document.

(6) Compare each text word with the Key Glossary and tag each Key word in text with a weight k_i (= frequency of occurrence).

(7) Compute the Key weight K of each sentence by summing its Key-word weights k_i.

(8) Rank all sentences in decreasing weight order.

(9) Select all sentences whose rank order is less than the threshold S(= 25 percent of the total number of sentences).

(10) Select all headings.

(11) Merge the selected sentences under their proper headings.

(12) Output title, authors, and the results of step (11).

2.2.3 Title Method

The Title method has as its source of machine-recognizable clues the specific characteristics of the skeleton of the document, i.e., title, headings, format, (see Section 3, Rationale of the Four Basic Methods). The Title method is based on a Title Glossary comprising those content words found in the title, subtitles, and headings of the document to be abstracted, but excluding words of the Null Dictionary (see Section 3, Rationale of the Four Word Lists).

The computation of sentence weights by the Title method is detailed in the following steps which also show the present choice of parameter values:

(1) Compare each word of title and headings with Null Dictionary.

(2) Create Title Glossary of nonmatching words.

(3) Tag all words of the Title Glossary with the weight t_1 (= 11) if they are words of the title.

Tag all words of the Title Glossary with the weight t_2 (= 7) if they are words of a heading.

(4) Compare each word of text with Title Glossary and tag each matching word with its Title weight t_1 or t_2.

(5) Compute the Title weight T of each sentence by summing its title word weights t_1 and t_2.

(6) Rank all sentences in decreasing weight order.

(7) Select all sentences whose rank order is less than the threshold S(= 25 percent of the total number of sentences).

(8) Select all headings.

(9) Merge the selected sentences under their proper headings.

(10) Output title, authors, and the results of step (9).

2.2.4 Location Method

The Location method has as its source of machine-recognizable clues the general characteristics of the corpus that are provided by the skeletons of documents, i.e., title, headings, format (see Section 3, Rationale of the Four Basic Methods). The Location method is based on a Heading Dictionary of certain function words that appear in skeletons of documents (see Section 3, Rationale of the Four Word Lists). Words of the Heading Dictionary together with their weights are presented in Exhibit 5. In addition to assigning weights provided by the Heading Dictionary, the Location method also assigns weights to sentences according to their ordinal position in the text, i.e., first and last paragraphs, and first and last sentences of paragraphs. The final Location weight for each sentence is the sum of the Heading weight and the Ordinal weight.

Computation of sentence weights by the Location method is detailed in the following steps which also show the present choice of parameter values:

(1) Compare each word of title and headings with Null Dictionary.

(2) Create table of nonmatching words.

(3) Compare table with Heading Dictionary.

(4) Tag each matching word with weight h_i found in Heading Dictionary.

(5) Compute the Heading weight H of each heading by summing its Heading-word weights h_i.

(6) Tag each sentence with the Heading weight H of its heading.

(7) Tag each sentence of the first paragraph with an Ordinal weight O_1 (= 18)

Tag each sentence of the last paragraph with an Ordinal weight O_2 (= 18)

Tag the first sentence of every paragraph with an Ordinal weight O_3 (= 9)

Tag the last sentence of every paragraph with an Ordinal weight O_4 (= 9)

(8) Compute the Ordinal weight O of each sentence by summing its ordinal weights O_i.

(9) Compute the Location weight L of each sentence by summing its Heading weight H and Ordinal weight O.

(10) Rank all sentences in decreasing weight order.

(11) Select all sentences whose rank order is less than the threshold S(= 25 percent of the total number of sentences).

(12) Select all headings.

(13) Merge the selected sentences under their proper headings.

(14) Output title, authors, and the results of step (13).

2.2.5 Combined Methods

The four basic methods described above yield 15 combined methods. They have in common the following sequence of final machine steps:

(1) Rank all sentences in decreasing weight order.

(2) Select the sentences whose rank order is less than the threshold S(= 25 percent of the total number of sentences).

(3) Select all headings.

(4) Merge the selected sentences under their proper headings.

(5) Output title, authors, and the results of step (4).

Heading Word	Weight	Heading Word	Weight
1. abstract	90	46. nomenclature	50
2. aim	90	47. object	90
3. aims	90	48. objective	90
4. analysis	10	49. objectives	90
5. approach	30	50. observations	10
6. background	50	51. performance	10
7. comparison	20	52. preliminary	20
8. concluding	90	53. problem	90
9. conclusion	90	54. problems	90
10. conclusions	90	55. program	30
11. consequence	90	56. progress	40
12. consequences	90	57. project	90
13. consideration	30	58. projects	90
14. considerations	30	59. properties	10
15. data	10	60. purpose	90
16. description	20	61. purposes	90
17. descriptions	20	62. recommendation	90
18. design	20	63. recommendations	90
19. determination	20	64. recommended	90
20. determinations	20	65. remarks	60
21. development	10	66. requirements	50
22. developments	10	67. results	30
23. discussion	30	68. review	40
24. effect	20	69. scope	50
25. effects	20	70. significance	90
26. evaluation	30	71. studies	10
27. evaluations	30	72. study	10
28. extension	90	73. subject	90
29. extensions	90	74. subjects	90
30. finding	90	75. suggested	90
31. findings	90	76. suggestion	90
32. foreword	20	77. suggestions	90
33. future	50	78. summary	90
34. generalization	90	79. symbols	20
35. generalizations	90	80. technical	10
36. goal	90	81. technique	50
37. goals	90	82. techniques	20
38. implication	90	83. test	10
39. implications	90	84. testing	10
40. introduction	60	85. tests	10
41. introductory	50	86. theoretical	30
42. investigation	30	87. theory	30
43. measurement	10	88. topic	90
44. method	50	89. topics	90
45. methods	20	90. work	20

Exhibit 5. Heading Dictionary

It would be expected that of these 15, the Cue-Key-Title-Location combined method would produce the best abstract, since all of the four categories of machine-recognizable clues come into play. However, the experimental data show that the combined Cue-Title-Location method, excluding the Key method, gives as good or better abstracts as the combination of all four methods. The decision to abandon the Key method is reinforced by the consideration that upon reprogramming considerable computer storage will be saved by omitting the Key routine. The experimental data underlying this decision are given in Section 3. The data support a conjecture made during the previous study that Key words, while important for indexing, are not as important for abstracting.

2.3 OUTPUT

Hardware

As in the case of input we are here confronted with unfortunate restrictions imposed by output hardware. Despite the fact that high speed printers are available, the most serious difficulty is that of an overly restricted number of type fonts. This forces us to replace strings of unusual symbols, e.g., mathematical and chemical, by the few conventional symbols available at the output printer. Moreover, we are forced to replace important strings of textual symbols by only one or two such conventional output symbols. This limitation has been taken into account in the sentence selection for the target abstracts by selecting only from the pre-edited document. The result was that if the significant content of a sentence could not be output by the computer it was discredited.

Format

The format of the classical abstract prepared by humans comprises title, author, and a paragraph of connected text. The present output hardware provides little leeway in the composition of the textual output. However, since an automatic abstract is, at present, nothing more than an automatic extract, it is desirable to correct the generally disjointed sequence of selected sentences by other devices. This problem has been partially solved by capturing in an automatic abstract those informative features of structure found in section headings and subheadings.

Dissemination

Even though the problem of dissemination of automatic abstracts has received little attention in the literature, it nevertheless will play an important part in the general acceptability and utility of automatic abstracts. Both theoretical and practical studies must be made to ascertain how the requester communicates with the abstracting system, how the system collates similar requests, and how the system produces multiple copies of the abstracts through a suitable medium.

2.3.1 System Output

An abstract of a document can be produced by any combination of the four basic methods of calculating the sentence weights in a document: Cue, Key, Title, and Location, which are defined in Section 2.2 of this report. The system output gives the following information:

(1) Document number

(2) Methods used to produce abstract

(3) Title of document

(4) Author of document

(5) Selected sentences (All are tagged with paragraph and sentence number.)

(6) All headings (Headings are identified by their 0 sentence number.)

Exhibits 6-11 show automatic abstracts produced by the six methods of greatest interest: C (Cue), K (Key), T (Title), L (Location), and the combined methods C-K-T-L, C-T-L.

Splitting of Abstracts

Depending on the order in which documents are placed on the Edit tape, it is possible that documents not over 3069 words can be split by the Abstracting Program into two abstracts. It is also possible that a document over 3069 words can be processed due to its arrangement on the Edit tape without a split occurring. These may never occur but the possibility of their happening does exist (see Appendix, Section 5).

2.3.2 Research Output

A. Key Word List:

Two lists of Key words are output. The first list is output in frequency order; an example is given in Exhibit 12. The second list is output in alphabetical order; an example is given in Exhibit 13. A description of how the Key words are computed is given in Section 2.2.2 of this report.

B. Sentence Weight List:

Two lists are output of each sentence number in the document along with its computed weight. The first list is in sentence order; an example is given in Exhibit 14. The second list is in weight order; an example is given in Exhibit 15. The weight is the sum resulting from the Cue, Key, Title, or Location methods used.

C. Abstract:

Defined in Section 2.3.1.

D. Vertical Listing:

This is a list of every word as it occurred in the document along with:

(1) Punctuation before (PB): includes open quotation mark, open parenthesis, initial capitalization (denoted by *), etc.

(2) 1st punctuation after (PA1): includes comma, hyphen, period, etc.

(3) 2nd punctuation after (PA2): includes hyphen, closed parenthesis, closed quotation mark, etc.

(4) Paragraph number (P)

(5) Sentence number within the paragraph (S)

(6) Word number within the sentence (W)

(7) Cue weight C (the weight computed according to Section 2. 2. 1; the Cue weight is only computed if switch 1 is set equal 1)

(8) Key weight K (the weight computed according to Section 2. 2. 2; occurs only if switch 3 is set equal 1)

(9) Title weight T (the weight computed according to Section 2. 2. 3; output only if switch 2 is set equal 1)

After the last word of each sentence is output, the Location weight for the sentence is output together with the total Cue, Key, and Title weights for that sentence. The Location weight is described in Section 2. 2. 4.

Exhibit 16 shows a portion of a vertical listing of Document 6 from the Exotic Fuel Corpus.

PAR SENT DOCUMENT NUMBER 6 PAGE 1

ABSTRACT BASED ON CUE WTS.

EVALUATION OF THE EFFECT CF DIMETHYLAMINE BORINE AND SEVERAL OTHER ADDITIVES ON COMBUSTION STABILITY CHARACTERISTICS OF VARIOUS HYDROCARBON TYPE FUELS IN PHILLIPS MICROBURNER (AD87730)

R. L. BRACE

PAR	SENT	
1	0	SUMMARY
3	2	PREVIOUS STUDIES IN PHILLIPS 2 INCH TURBOJET ENGINE TYPE COMBUSTOR HAD INDICATED THAT SUCH MATERIALS COULD SUBSTANTIALLY INCREASE THE MAXIMUM RATE OF HEAT RELEASE ATTAINABLE, ESPECIALLY WITH LOW PERFORMANCE FUELS SUCH AS THE ISO PARAFFIN TYPE HYDROCARBONS-PARTICULARLY WHEN OPERATING UNDER SEVERE CCNDITICNS FOR COMBUSTION (I.E. , HIGH AIR FLOW VELOCITY OR LOW COMBUSTION PRESSURE).
4	1	THE ASSUMPTICN HAS BEEN MACE IN THIS FUEL EVALUATION THAT THE GREATER THE ALLOWABLE HEAT INPUT RATE FOR A GIVEN VELOCITY, THE GREATER THE DEGKEE OF COMBUSTION STABILITY.
4	2	CN THIS BASIS, THE DATA INDICATE THAT ALL THE ADDITIVE MATERIALS TESTED CAUSED AN INCREASE IN STABILITY PERFORMANCE., A FUEL OF RELATIVELY LOW PERFORMANCE SUCH AS TOLUENE BEING BENEFITED TO A GREATER EXTENT THAN A HIGH PERFORMANCE FUEL SUCH AS NORMAL HEPTANE.
4	5	IN GENERAL, ADDITIVE CONCENTRATIONS OF ONE PER CENT BY WEIGHT IN THE SEVERAL PURE HYDROCARBONS WHICH NORMALLY DIFFERED QUITE WIDELY IN PERFORMANCE, PRODUCED UNIFORMLY SUPERIOR COMBUSTION STABILITY CHARACTERISTICS AS MEASURED USING THE PHILLIPS MICROBURNER.
5	0	I. INTROCUCTION
6	1	AT THE REQUEST OF THE NAVY BUREAU OF AERONAUTICS THE JET FUELS GROUP HAS EVALUATED THE EFFECTS OF THE ADDITION OF SMALL AMOUNTS OF DIMETHYLAMINEBORIN =S ON THE COMBUSTION STABILITY PERFORMANCE OF SEVERAL HYDROCARBON FUELS.
8	0	II. DESCRIPTION OF PHILLIPS MICROBURNER (MODEL 1A)
10	0	III. DESCRIPTION OF TEST APPARATUS
11	2	IN THE PRESENT EVALUATION IT WAS NOT NECESSARY TO CONSIDER THE EFFECT OF CORROSION, CONSEQUENTLY A CONTINUOUS FLOW SYSTEM PROVIDING GREATER FLEXIBILITY AND EASIER HANDLING WAS INCORPORATED WHICH REQUIRES ONLY SLIGHTLY MORE FUEL PER TEST THAN THE ORIGINAL.
14	0	IV. DESCRIPTICN OF TEST FUELS
15	2	THESE FUELS REPRESENT VARIATIONS IN CHEMICAL STRUCTURE WHICH WILL IN TURN PROVIDE INDICES OF BOTH GOOD AND POOR COMBUSTION STABILITY PERFORMANCE.
17	0	V. TEST PROCEDURE
21	0	VI. RESULTS
22	2	THE RECORDED DATA WERE CONVERTED TO VALUES OF HEAT INPUT RATE AND A REFERENCE VELOCITY AT FLASHBACK-THUS THE FLASHBACK LIMITS (OR COMBUSTION STABILITY CHARACTERISTICS) ARE ESTABLISHED ON THE BASIS OF AN ALLOWABLE HEAT INPUT RATE.
24	1	THE REGICN OF STABLE OPERATION IS DEFINED AS THE STATE OF FLASHBACKTHE CONDITIONS OF COMBUSTION WHERE THE FLAME WOULD BECOME ANCHORED TO A FLAME HOLDER-AS IN STABLE GAS TURBINE OR RAM JET COMBUSTOR OPERATION-IF THE FLAME HOLDER WERE PROVIDED IN THE BURNER TUBE.
25	0	VII. DISCUSSION
29	1	PREVIOUS WORK CONDUCTED IN THE PHILLIPS 2 INCH COMBUSTOR (REF. 2) INDICATED THAT SOME ADDITIVES CAUSED A SIGNIFICANT INCREASE IN THE PERFORMANCE OF A LOW RATING FUEL WHILE THESE SAME ADDITIVES DID NOT SUBSTANTIALLY EFFECT THE HIGHER RATING FUELS.
30	3	THE •BOOST• EFFECTED WITH ADDITIVE IN TOLUENE IS ABOUT FOUR TIMES THAT OF THE INCREASE AFFORDED WITH BENZENE, INDICATING THE PHENOMANA TO BE ONE OF GENERAL PERFORMANCE LEVEL RATHER THAN A CHARACTERISTIC OF AROMATIC TYPE HYDROCARBONS.
31	2	EXAMINATION OF THESE CURVES SHOWS DIMETHYLAMINE BORINE TO PROVIDE THE GREATEST INCREASE IN STABILITY PERFORMANCE AND PROPYLENE OXIDE THE LEAST.
31	3	ALL FOUR ADDITIVES INDICATED THEIR ADDITION TO BE SUBJECT TO THE EFFECT OF DEMINISHING RESULTS UPON FURTHER ADDITION-THAT IS, THEIR EFFECT WAS NOT ESSENTIALLY A BLENDING EFFECT.
33	0	VIII. CONCLUSIONS
34	1	1. THE ACDITION OF DIMETHYLAMINE BORINE IN CONCENTRATIONS OF ONE PER CENT BY WEIGHT TO JET FUEL TYPE HYDROCARBONS RESULTED IN A UNIFORMLY HIGH LEVEL OF COMBUSTION STABILITY PERFORMANCE AS MEASURED BY PHILLIPS MICROBURNER.
35	1	2. THE ACDITION OF RELATIVELY LARGE AMOUNTS OF PROPYLENE OXIDE TO TOLUENE WERE NECESSARY TO PROVIDE SIGNIFICANT IMPROVEMENT IN STABILITY PERFORMANCE AS INDICATEC BY INCREASES IN ALLOWABLE HEAT INPUT RATES.
36	1	3. THE ADDITION OF ADDITIVE CCNCENTRATIONS (UP TO 1 PER CENT) OF AMYL NITRATE, CUMENE HYDROPEROXIDE, AND DIMETHYLAMINE BORINE ALL RESULTED IN IMPROVED STABILITY PERFORMANCE., THE GREATEST INCREASES WERE SHOWN WHEN BLENDED WITH A FUEL OF POOR PERFORMANCE CHARACTERISTICS-SUCH AS TOLUENE.
36	2	BENEFICIAL EFFECTS WERE APPRECIABLY LESS WHEN BLENDED WITH A FUEL OF GOOD PERFORMANCE CHARACTERISTICS-SUCH AS N-HEPTANE.
37	0	IX. RECOMMENDATIONS
38	1	BASED ON THE EVALUATION OF THE EFFECTS OF ADDITIVES ON THE FLASHBACK LIMITS OF THE ADDITIVE-FUEL BLENDS TESTED IN THE MICROBURNER (MODEL 1A) IT IS RECOMMENDED THAT DIMETHYLAMINE BORINE SHOULD BE FURTHER INVESTIGATED.
38	2	THIS FUTURE WORK SHOULD INCLUDE STUDY OF COMBUSTION STABILITY AND COMBUSTION EFFICIENCY EFFECTS IN THE PHILLIPS 2 INCH COMBUSTOR AND AN INVESTIGATION OF ITS INFLUENCE ON COMBUSTION CLEANLINESS.

Exhibit 6. Automatic Abstract of Document 6 Produced by Cue Method C

PAR SENT DOCUMENT NUMBER 6 PAGE 1

ABSTRACT BASED ON KEY WTS.

EVALUATION OF THE EFFECT CF DIMETHYLAMINE BORINE AND SEVERAL OTHER ADDITIVES ON COMBUSTION STABILITY CHARACTERISTICS OF VARIOUS HYDROCARBON TYPE FUELS IN PHILLIPS MICROBURNER (AD87730)

R. L. BRACE

PAR	SENT	
1	0	SUMMARY
2	1	AT THE REQUEST OF THE NAVY BUREAU OF AERONAUTICS, PHILLIPS PETROLEUM COMPANY UNDERTOOK THE EVALUATION OF DIMETHYLAMINE BORINE AS AN ADDITIVE FOR IMPROVING THE COMBUSTION CHARACTERISTICS OF AVIATION GAS TURBINE TYPE FUELS.
3	2	PREVIOUS STUDIES IN PHILLIPS 2 INCH TURBOJET ENGINE TYPE COMBUSTOR HAD INDICATED THAT SUCH MATERIALS COULD SUBSTANTIALLY INCREASE THE MAXIMUM RATE OF HEAT RELEASE ATTAINABLE, ESPECIALLY WITH LOW PERFORMANCE FUELS SUCH AS THE ISO PARAFFIN TYPE HYDROCARBONS-PARTICULARLY WHEN OPERATING UNDER SEVERE CONDITIONS FOR COMBUSTION (I.E. , HIGH AIR FLOW VELOCITY OR LOW COMBUSTION PRESSURE).
4	1	THE ASSUMPTION HAS BEEN MADE IN THIS FUEL EVALUATION THAT THE GREATER THE ALLOWABLE HEAT INPUT RATE FOR A GIVEN VELOCITY, THE GREATER THE DEGREE OF COMBUSTION STABILITY.
4	2	ON THIS BASIS, THE DATA INDICATE THAT ALL THE ADDITIVE MATERIALS TESTED CAUSED AN INCREASE IN STABILITY PERFORMANCE., A FUEL OF RELATIVELY LOW PERFORMANCE SUCH AS TOLUENE BEING BENEFITED TO A GREATER EXTENT THAN A HIGH PERFORMANCE FUEL SUCH AS NORMAL HEPTANE.
4	4	WITH RESPECT TO THE DIMETHYLAMINE BORINE, ITS EFFECT AS A FUEL ADDITIVE WAS NOTEWORTHY., 0.1 WEIGHT PER CENT IN TOLUENE BEING EQUIVALENT TO 20 PER CENT BY WEIGHT OF ADDED PROPYLENE OXIDE.
4	5	IN GENERAL, ADDITIVE CONCENTRATIONS OF ONE PER CENT BY WEIGHT IN THE SEVERAL PURE HYDROCARBONS WHICH NORMALLY DIFFERED QUITE WIDELY IN PERFORMANCE, PRODUCED UNIFORMLY SUPERIOR COMBUSTION STABILITY CHARACTERISTICS AS MEASURED USING THE PHILLIPS MICROBURNER.
5	0	I. INTRODUCTION
8	0	II. DESCRIPTION OF PHILLIPS MICROBURNER (MODEL 1A)
10	0	III. DESCRIPTION OF TEST APPARATUS
14	0	IV. DESCRIPTION OF TEST FUELS
17	0	V. TEST PROCEDURE
21	0	VI. RESULTS
22	3	THE HEAT INPUT RATE IS DETERMINED BY THE HEATING VALUE OF THE FUEL PER POUND OF AIR OR HIR-LOWER HEATING VALUE X (FUEL FLOW/AIR FLOW) FB
25	0	VII. DISCUSSION
26	1	THE EVALUATION OF THE EFFECTS OF THE ADDITIVES UNDER CONSIDERATION ON THE COMBUSTION STABILITY CHARACTERISTICS OF THE FUEL BLENDS TESTED WERE INTERPRETED FROM A PLOT OF HEAT INPUT RATE AT FLASHBACK AGAINST REFERENCE VELOCITY.
29	2	THEREFORE, IN ADDITION TO THE EVALUATION OF THE FUEL BLENDS CONTAINING 0.1, 0.5, AND 1 PER CENT DIMETHYLAMINE BORINE IN TOLUENE (SHOWN IN FIGURE 5) THE SAME ADDITIVE WAS TESTED IN N-HEPTANE (SHOWN IN FIGURE 6).
31	1	IN ORDER TO COMPARE THE ADDITIVES EVALUATED, THE ALLOWABLE HEAT INPUT RATES OF THE FUEL-ADDITIVE BLENDS AT A CONSTANT VELOCITY OF 12 FPS (ARBITRARILY SELECTED) WERE PLOTTED AGAINST ADDITIVE CONCENTRATION-SHOWN IN FIGURE 9.
32	1	MENTION SHOULD BE MADE OF THE FACT THAT DURING THE COMBUSTION OF THE DIMETHYLAMINE BORINE-HYDROCARBON FUEL BLENDS NO NOTICEABLE ODORS OR SMOKE WERE OBSERVED.
33	0	VIII. CONCLUSIONS
34	1	1. THE ADDITION OF DIMETHYLAMINE BORINE IN CONCENTRATIONS OF ONE PER CENT BY WEIGHT TO JET FUEL TYPE HYDROCARBONS RESULTED IN A UNIFORMLY HIGH LEVEL OF COMBUSTION STABILITY PF ORMANCE AS MEASURED BY PHILLIPS MICROBURNER.
36	1	3. THE ADDITION OF ADDITIVE CONCENTRATIONS (UP TO 1 PER CENT) OF AMYL NITRATE, CUMENE HYDROPEROXIDE, AND DIMETHYLAMINE BORINE ALL RESULTED IN IMPROVED STABILITY PERFORMANCE., THE GREATEST INCREASES WERE SHOWN WHEN BLENDED WITH A FUEL OF POOR PERFORMANCE CHARACTERISTICS-SUCH AS TOLUENE.
37	0	IX. RECOMMENDATIONS
38	1	BASED ON THE EVALUATION OF THE EFFECTS OF ADDITIVES ON THE FLASHBACK LIMITS OF THE ADDITIVE-FUEL BLENDS TESTED IN THE MICROBURNER (MODEL 1A) IT IS RECOMMENDED THAT DIMETHYLAMINE BORINE SHOULD BE FURTHER INVESTIGATED.
38	2	THIS FUTURE WORK SHOULD INCLUDE STUDY OF COMBUSTION STABILITY AND COMBUSTION EFFICIENCY EFFECTS IN THE PHILLIPS 2 INCH COMBUSTOR AND AN INVESTIGATION OF ITS INFLUENCE ON COMBUSTION CLEANLINESS.

Exhibit 7. Automatic Abstract of Document 6 Produced by Key Method K

PAR SENT DOCUMENT NUMBER 6 PAGE 1

ABSTRACT BASED ON TITLE WTS.

EVALUATION OF THE EFFECT OF DIMETHYLAMINE BORINE ANC SEVERAL OTHER ADDITIVES ON COMBUSTION STABILITY CHARACTERISTICS OF VARIOUS HYDROCARBON TYPE FUELS IN PHILLIPS MICROBURNER (AD87730)

R. L. BRACE

PAR	SENT	
1	0	SUMMARY
2	1	AT THE REQUEST OF THE NAVY BUREAU OF AERCNAUTICS, PHILLIPS PETROLEUM COMPANY UNDERTOOK THE EVALUATION OF DIMETHYLAMINE BORINE AS AN ADDITIVE FOR IMPROVING THE COMBUSTION CHARACTERISTICS OF AVIATION GAS TURBINE TYPE FUELS.
2	2	BECAUSE OF THE SMALL AMOUNT (100 GRAMS) OF DIMETHYLAMINE RORINE RECEIVED FROM CALLERY CHEMICAL COMPANY, THIS EVALUATION HAS BEEN LIMITED TO THE MEASURCMENT OF ITS EFFECT ON THE FLASH-BACK CHARACTERISTICS OF THREE PURE HYDROCARBCNS (TOLUENE, NORMAL HEPTANE AND RENZENE) IN THE PHILLIPS MICROBURNER.
3	2	PREVIOUS STUDIES IN PHILLIPS 2 INCH TURBOJET ENGINE TYPE COMBUSTOR HAD INDICATED THAT SUCH MATERIALS COULD SUBSTANTIALLY INCREASE THE MAXIMUM RATE OF HEAT RELEASE ATTAINABLE, ESPECIALLY WITH LOW PERFORMANCE FUELS SUCH AS THE ISO PARAFFIN TYPE HYDROCARBONS-PARTICULARLY WHEN OPERATING UNDER SEVERE CONDITIONS FOR COMBUSTION (I.E. , HIGH AIR FLOW VELOCITY OR LOW COMBUSTION PRESSURE).
4	5	IN GENERAL, ADDITIVE CONCENTRATIONS OF CNE PER CENT BY WEIGHT IN THE SEVERAL PURE HYDROCARBONS WHICH NORMALLY DIFFERED QUITE WIDELY IN PERFORMANCE, PRODUCED UNIFORMLY SUPERIOR COMBUSTION STABILITY CHARACTERISTICS AS MEASURED USING THE PHILLIPS MICROBURNER.
5	0	I. INTRODUCTION
6	1	AT THE REQUEST OF THE NAVY BUREAU OF AERCNAUTICS THE JET FUELS GROUP HAS EVALUATED THE EFFECTS OF THE ADDITION OF SMALL AMOUNTS OF DIMETHYLAMINEBORIN =S ON THE COMBUSTION STABILITY PERFORMANCE OF SEVERAL HYDROCARBON FUELS.
7	1	DUE TO THE SMALL QUANTITY OF THIS MATERIAL OBTAINED THE EVALUATION WAS CONDUCTED IN THE PHILLIPS MICROBURNER (MODEL 1A) WHICH IS A SLIGHTLY MODIFIED VERSION OF THE ORIGINAL PHILLIPS MICROBURNER (MODEL 1).
8	0	II. DESCRIPTION OF PHILLIPS MICROBURNER (MODEL 1A)
10	0	III. DESCRIPTION OF TEST APPARATUS
14	0	IV. DESCRIPTION OF TEST FUELS
16	1	THE ADDITIVES EVALUATED, DIMETHYLAMINE BORINE, PROPYLENE OXIDE, AMYL NITRATE, AND CUMENE HYDROPAROXIDE, WERE BLENDED INTO THE HYDROCARBON FUEL RY WEIGHT IN CONCENTRATIONS RANGING FROM 0.1 TO 20 PER CENT.
17	0	V. TEST PROCEDURE
21	0	VI. RESULTS
25	0	VII. DISCUSSION
26	1	THE EVALUATION OF THE EFFECTS OF THE ADDITIVES UNDER CONSIDERATION CN THE COMBUSTION STABILITY CHARACTERISTICS OF THE FUEL BLENDS TESTED WERE INTERPRETED FROM A PLOT OF HEAT INPUT RATE AT FLASHBACK AGAINST REFERENCE VELOCITY.
29	1	PREVIOUS WORK CONDUCTED IN THE PHILLIPS 2 INCH COMBUSTOR (REF. 2) INDICATED THAT SOME ADDITIVES CAUSED A SIGNIFICANT INCREASE IN THE PERFORMANCE OF A LOW RATING FUEL WHILE THESE SAME ADDITIVES DID NOT SUBSTANTIALLY EFFECT THE HIGHER RATING FUELS.
30	1	TO MORE DEFINITELY ESTABLISH THE EFFECT OF DIMETHYLAMINE BORINE ON THE FLASHBACK LIMITS OF AROMATIC TYPE HYDROCARBON FUELS, BENZENE WITH 0.84 WEIGHT PER CENT ADDITIVE WAS EVALUATED AS SHOWN IN FIGURE 8.
31	3	ALL FOUR ADDITIVES INDICATED THEIR ADDITION TO BE SUBJECT TO THE EFFECT OF DEMINISHING RESULTS UPON FURTHER ADDITION-THAT IS, THEIR EFFECT WAS NOT ESSENTIALLY A BLENDING EFFECT.
32	1	MENTION SHOULD BE MADE OF THE FACT THAT DURING THE COMBUSTION OF THE DIMETHYLAMINE BORINE-HYDROCARBON FUEL BLENDS NO NOTICEABLE ODORS OR SMOKE WERE OBSERVED.
33	0	VIII. CONCLUSIONS
34	1	1. THE ADDITION OF DIMETHYLAMINE BORINE IN CONCENTRATIONS OF ONE PER CENT BY WEIGHT TO JET FUEL TYPE HYDROCARBONS RESULTED IN A UNIFORMLY HIGH LEVEL OF COMBUSTION STARILITY PERFORMANCE AS MEASURED BY PHILLIPS MICROBURNER.
36	1	3. THE ADDITION OF ADDITIVE CONCENTRATICNS (UP TO 1 PER CENT) OF AMYL NITRATE, CUMENE HYDROPEROXIDE, AND DIMETHYLAMINE BORINE ALL RESULTED IN IMPROVED STABILITY PERFORMANCE., THE GREATEST INCREASES
		WERE SHOWN WHEN BLENDED WITH A FUEL OF PCOR PERFORMANCE CHARACTERISTICS-SUCH AS TOLUENE.
37	0	IX. RECOMMENDATIONS
38	1	BASED ON THE EVALUATION OF THE EFFECTS OF ADDITIVES ON THE FLASHBACK LIMITS OF THE ADDITIVE-FUEL BLENDS TESTED IN THE MICROBURNER (MODEL 1A) IT IS RECOMMENDED THAT DIMETHYLAMINE BORINE SHOULD BE FURTHER INVESTIGATED.
38	2	THIS FUTURE WORK SHOULD INCLUDE STUDY OF COMBUSTION STABILITY AND COMBUSTION EFFICIENCY EFFECTS IN THE PHILLIPS 2 INCH COMBUSTOR AND AN INVESTIGATION OF ITS INFLUENCE ON COMBUSTION CLEANLINESS.

Exhibit 8. Automatic Abstract of Document 6 Produced by Title Method T

PAR SENT DOCUMENT NUMBER 6 PAGE 1

ABSTRACT BASED ON LOC. WTS.

EVALUATION OF THE EFFECT OF DIMETHYLAMINE BORINE AND SEVERAL OTHER ADDITIVES ON COMBUSTION STABILITY CHARACTERISTICS OF VARIOUS HYDROCARBON TYPE FUELS IN PHILLIPS MICROBURNER (AD87730)

R. L. BRACE

PAR	SENT	
1	0	SUMMARY
2	1	AT THE REQUEST OF THE NAVY BUREAU OF AERONAUTICS, PHILLIPS PETROLEUM COMPANY UNDERTOOK THE EVALUATION OF DIMETHYLAMINE BORINE AS AN ADDITIVE FOR IMPROVING THE COMBUSTION CHARACTERISTICS OF AVIATION GAS TURBINE TYPE FUELS.
2	2	BECAUSE OF THE SMALL AMOUNT (100 GRAMS) OF DIMETHYLAMINE BORINE RECEIVED FROM CALLERY CHEMICAL COMPANY, THIS EVALUATION HAS BEEN LIMITED TO THE MEASUREMENT OF ITS EFFECT ON THE FLASH-BACK CHARACTERISTICS OF THREE PURE HYDROCARBONS (TOLUENE, NORMAL HEPTANE AND BENZENE) IN THE PHILLIPS MICROBURNER.
2	3	DIMETHYLAMINE BORINE CONCENTRATIONS OF FROM 0.1 TO 1.0 PER CENT BY WEIGHT WERE EVALUATED.
3	1	FOR COMPARATIVE PURPOSES TWO COMMON IGNITION ADDITIVES (AMYL NITRATE AND CUMENE HYDROPEROXIDE) WERE ALSO EVALUATED DURING THIS STUDY, AS WELL AS CONCENTRATIONS UP TO 20 PER CENT BY WEIGHT OF PROPYLENE OXIDE-A RELATIVELY HIGH FLAME VELOCITY FUEL.
3	2	PREVIOUS STUDIES IN PHILLIPS 2 INCH TURBOJET ENGINE TYPE COMBUSTOR HAD INDICATED THAT SUCH MATERIALS COULD SUBSTANTIALLY INCREASE THE MAXIMUM RATE OF HEAT RELEASE ATTAINABLE, ESPECIALLY WITH LOW PERFORMANCE FUELS SUCH AS THE ISO PARAFFIN TYPE HYDROCARBONS-PARTICULARLY WHEN OPERATING UNDER SEVERE CONDITIONS FOR COMBUSTION (I.E. , HIGH AIR FLOW VELOCITY OR LOW COMBUSTION PRESSURE).
4	1	THE ASSUMPTION HAS BEEN MADE IN THIS FUEL EVALUATION THAT THE GREATER THE ALLOWABLE HEAT INPUT RATE FOR A GIVEN VELOCITY, THE GREATER THE DEGREE OF COMBUSTION STABILITY.
4	2	ON THIS BASIS, THE DATA INDICATE THAT ALL THE ADDITIVE MATERIALS TESTED CAUSED AN INCREASE IN STABILITY PERFORMANCE., A FUEL OF RELATIVELY LOW PERFORMANCE SUCH AS TOLUENE BEING BENEFITED TO A GREATER EXTENT THAN A HIGH PERFORMANCE FUEL SUCH AS NORMAL HEPTANE.
4	3	THESE DATA ARE IN AGREEMENT WITH PREVIOUS ADDITIVE STUDIES BY PHILLIPS.
4	4	WITH RESPECT TO THE DIMETHYLAMINE BORINE, ITS EFFECT AS A FUEL ADDITIVE WAS NOTEWORTHY., 0.1 WEIGHT PER CENT IN TOLUENE BEING EQUIVALENT TO 20 PER CENT BY WEIGHT OF ADDED PROPYLENE OXIDE.
4	5	IN GENERAL, ADDITIVE CONCENTRATIONS OF ONE PER CENT BY WEIGHT IN THE SEVERAL PURE HYDROCARBONS WHICH NORMALLY DIFFERED QUITE WIDELY IN PERFORMANCE, PRODUCED UNIFORMLY SUPERIOR COMBUSTION STABILITY CHARACTERISTICS AS MEASURED USING THE PHILLIPS MICROBURNER.
5	0	I. INTRODUCTION
8	0	II. DESCRIPTION OF PHILLIPS MICROBURNER (MODEL 1A)
10	0	III. DESCRIPTION OF TEST APPARATUS
14	0	IV. DESCRIPTION OF TEST FUELS
17	0	V. TEST PROCEDURE
21	0	VI. RESULTS
25	0	VII. DISCUSSION
33	0	VIII. CONCLUSIONS
34	1	1. THE ADDITION OF DIMETHYLAMINE BORINE IN CONCENTRATIONS OF ONE PER CENT BY WEIGHT TO JET FUEL TYPE HYDROCARBONS RESULTED IN A UNIFORMLY HIGH LEVEL OF COMBUSTION STABILITY PERFORMANCE AS MEASURED BY PHILLIPS MICROBURNER.
35	1	2. THE ADDITION OF RELATIVELY LARGE AMOUNTS OF PROPYLENE OXIDE TO TOLUENE WERE NECESSARY TO PROVIDE SIGNIFICANT IMPROVEMENT IN STABILITY PERFORMANCE AS INDICATED BY INCREASES IN ALLOWABLE HEAT INPUT RATES.
36	1	3. THE ADDITION OF ADDITIVE CONCENTRATIONS (UP TO 1 PER CENT) OF AMYL NITRATE, CUMENE HYDROPEROXIDE, AND DIMETHYLAMINE BORINE ALL RESULTED IN IMPROVED STABILITY PERFORMANCE., THE GREATEST INCREASES WERE SHOWN WHEN BLENDED WITH A FUEL OF POOR PERFORMANCE CHARACTERISTICS-SUCH AS TOLUENE.
36	2	BENEFICIAL EFFECTS WERE APPRECIABLY LESS WHEN BLENDED WITH A FUEL OF GOOD PERFORMANCE CHARACTERISTICS-SUCH AS N-HEPTANE.
37	0	IX. RECOMMENDATIONS
38	1	BASED ON THE EVALUATION OF THE EFFECTS OF ADDITIVES ON THE FLASHBACK LIMITS OF THE ADDITIVE-FUEL BLENDS TESTED IN THE MICROBURNER (MODEL 1A) IT IS RECOMMENDED THAT DIMETHYLAMINE BORINE SHOULD BE FURTHER INVESTIGATED.
38	2	THIS FUTURE WORK SHOULD INCLUDE STUDY OF COMBUSTION STABILITY AND COMBUSTION EFFICIENCY EFFECTS IN THE PHILLIPS 2 INCH COMBUSTOR AND AN INVESTIGATION OF ITS INFLUENCE ON COMBUSTION CLEANLINESS.

Exhibit 9. Automatic Abstract of Document 6 Produced by Location Method L

PAR SENT DOCUMENT NUMBER 6 PAGE 1

ABSTRACT BASED ON CUE KEY TITLE LOC. WTS.

EVALUATION OF THE EFFECT OF DIMETHYLAMINE BORINE AND SEVERAL OTHER ADDITIVES ON COMBUSTION STABILITY CHARACTERISTICS OF VARIOUS HYDROCARBON TYPE FUELS IN PHILLIPS MICROBURNER (AD87730)

R. L. BRACE

PAR	SENT	
1	0	SUMMARY
2	1	AT THE REQUEST OF THE NAVY BUREAU OF AERONAUTICS, PHILLIPS PETROLEUM COMPANY UNDERTOOK THE EVALUATION OF DIMETHYLAMINE BORINE AS AN ADDITIVE FOR IMPROVING THE COMBUSTION CHARACTERISTICS OF AVIATION GAS TURBINE TYPE FUELS.
2	2	BECAUSE OF THE SMALL AMOUNT (100 GRAMS) OF DIMETHYLAMINE BORINE RECEIVED FROM CALLERY CHEMICAL COMPANY, THIS EVALUATION HAS BEEN LIMITED TO THE MEASUREMENT OF ITS EFFECT ON THE FLASH-BACK CHARACTERISTICS OF THREE PURE HYDROCARBONS (TOLUENE, NORMAL HEPTANE AND BENZENE) IN THE PHILLIPS MICROBURNER.
3	2	PREVIOUS STUDIES IN PHILLIPS 2 INCH TURBOJET ENGINE TYPE COMBUSTOR HAD INDICATED THAT SUCH MATERIALS COULD SUBSTANTIALLY INCREASE THE MAXIMUM RATE OF HEAT RELEASE ATTAINABLE, ESPECIALLY WITH LOW PERFORMANCE FUELS SUCH AS THE ISO PARAFFIN TYPE HYDROCARBONS-PARTICULARLY WHEN OPERATING UNDER SEVERE CONDITIONS FOR COMBUSTION (I.E. , HIGH AIR FLOW VELOCITY OR LOW COMBUSTION PRESSURE).
4	1	THE ASSUMPTICN HAS BEEN MADE IN THIS FUEL EVALUATION THAT THE GREATER THE ALLOWABLE HEAT INPUT RATE FOR A GIVEN VELOCITY, THE GREATER THE DEGREE OF COMBUSTION STABILITY.
4	2	CN THIS BASIS, THE DATA INDICATE THAT ALL THE ADDITIVE MATERIALS TESTED CAUSED AN INCREASE IN STABILITY PERFORMANCE., A FUEL OF RELATIVELY LOW PERFORMANCE SUCH AS TOLUENE BEING BENEFITED TO A GREATER EXTENT THAN A HIGH PERFORMANCE FUEL SUCH AS NORMAL HEPTANE.
4	4	WITH RESPECT TO THE DIMETHYLAMINE BORINE, ITS EFFECT AS A FUEL ADDITIVE WAS NOTEWORTHY., 0.1 WEIGHT PER CENT IN TOLUENE BEING EQUIVALENT TO 20 PER CENT BY WEIGHT OF ADDED PROPYLENE OXIDE.
4	5	IN GENERAL, ADDITIVE CONCENTRATIONS OF ONE PER CENT BY WEIGHT IN THE SEVERAL PURE HYDROCARBONS WHICH NORMALLY DIFFERED QUITE WIDELY IN PERFORMANCE, PRODUCED UNIFORMLY SUPERIOR COMBUSTION STABILITY CHARACTERISTICS AS MEASURED USING THE PHILLIPS MICROBURNER.
5	0	I. INTRODUCTION
6	1	AT THE REQUEST OF THE NAVY BUREAU OF AERONAUTICS THE JET FUELS GROUP HAS EVALUATED THE EFFECTS OF THE ADDITION OF SMALL AMOUNTS OF DIMETHYLAMINEBORIN =S ON THE COMBUSTION STABILITY PERFORMANCE OF SEVERAL HYDROCARBON FUELS.
7	1	CUE TO THE SMALL QUANTITY OF THIS MATERIAL OBTAINED THE EVALUATION WAS CONDUCTED IN THE PHILLIPS MICROBURNER (MODEL 1A) WHICH IS A SLIGHTLY MODIFIED VERSION OF THE ORIGINAL PHILLIPS MICROBURNER (MODEL 1).
8	0	II. DESCRIPTION OF PHILLIPS MICROBURNER (MODEL 1A)
10	0	III. DESCRIPTION OF TEST APPARATUS
14	0	IV. DESCRIPTION OF TEST FUELS
17	0	V. TEST PROCEDURE
21	0	VI. RESULTS
25	0	VII. DISCUSSION
29	1	PREVIOUS WORK CONDUCTED IN THE PHILLIPS 2 INCH COMBUSTOR (REF. 2) INDICATED THAT SOME ADDITIVES CAUSED A SIGNIFICANT INCREASE IN THE PERFORMANCE OF A LOW RATING FUEL WHILE THESE SAME ADDITIVES DID NOT SUBSTANTIALLY EFFECT THE HIGHER RATING FUELS.
33	0	VIII. CONCLUSIONS
34	1	1. THE ADDITION OF DIMETHYLAMINE BORINE IN CONCENTRATIONS OF ONE PER CENT BY WEIGHT TO JET FUEL TYPE HYDROCARBONS RESULTED IN A UNIFORMLY HIGH LEVEL OF COMBUSTION STABILITY PERFORMANCE AS MEASURED BY PHILLIPS MICROBURNER.
35	1	2. THE ADDITION OF RELATIVELY LARGE AMOUNTS OF PROPYLENE OXIDE TO TOLUENE WERE NECESSARY TO PROVIDE SIGNIFICANT IMPROVEMENT IN STABILITY PERFORMANCE AS INDICATED BY INCREASES IN ALLOWABLE HEAT INPUT RATES.
36	1	3. THE ADDITION OF ADDITIVE CONCENTRATIONS (UP TO 1 PER CENT) OF AMYL NITRATE, CUMENE HYDROPEROXIDE, AND DIMETHYLAMINE BORINE ALL RESULTED IN IMPROVED STABILITY PERFORMANCE., THE GREATEST INCREASES WERE SHOWN WHEN BLENDED WITH A FUEL OF POOR PERFORMANCE CHARACTERISTICS-SUCH AS TOLUENE.
37	0	IX. RECOMMENDATIONS

DOCUMENT NUMBER 6 PAGE 3

PAR	SENT	
38	1	BASED ON THE EVALUATION OF THE EFFECTS OF ADDITIVES ON THE FLASHBACK LIMITS OF THE ADDITIVE-FUEL BLENDS TESTED IN THE MICROBURNER (MODEL 1A) IT IS RECOMMENDED THAT DIMETHYLAMINE BORINE SHOULD BE FURTHER INVESTIGATED.
38	2	THIS FUTURE WORK SHOULD INCLUDE STUDY OF COMBUSTION STABILITY AND COMBUSTION EFFICIENCY EFFECTS IN THE PHILLIPS 2 INCH COMBUSTOR AND AN INVESTIGATION OF ITS INFLUENCE ON COMBUSTION CLEANLINESS.

Exhibit 10. Automatic Abstract of Document 6 Produced by Combined Method C-K-T-L

PAR SENT DOCUMENT NUMBER 6 PAGE 1

ABSTRACT BASED ON CUE TITLE LOC. WTS.

EVALUATION OF THE EFFECT OF DIMETHYLAMINE BORINE AND SEVERAL OTHER ADDITIVES ON COMBUSTION STABILITY CHARACTERISTICS OF VARIOUS HYDROCARBON TYPE FUELS IN PHILLIPS MICROBURNER (AD87730)

R. L. BRACE

1	0	SUMMARY
2	1	AT THE REQUEST OF THE NAVY BUREAU OF AERONAUTICS, PHILLIPS PETROLEUM COMPANY UNDERTOOK THE EVALUATION OF DIMETHYLAMINE BORINE AS AN ADDITIVE FOR IMPROVING THE COMBUSTION CHARACTERISTICS OF AVIATION GAS TURBINE TYPE FUELS.
2	2	BECAUSE OF THE SMALL AMOUNT (100 GRAMS) OF DIMETHYLAMINE BORINE RECEIVED FROM CALLERY CHEMICAL COMPANY, THIS EVALUATION HAS BEEN LIMITED TO THE MEASUREMENT OF ITS EFFECT ON THE FLASH-BACK CHARACTERISTICS OF THREE PURE HYDROCARBCNS (TOLUENE, NORMAL HEPTANE AND BENZENE) IN THE PHILLIPS MICROBURNER.
3	2	PREVIOUS STUDIES IN PHILLIPS 2 INCH TURBOJET ENGINE TYPE COMBUSTOR HAD INDICATED THAT SUCH MATERIALS COULD SUBSTANTIALLY INCREASE THE MAXIMUM RATE OF HEAT RELEASE ATTAINABLE, ESPECIALLY WITH LOW PERFORMANCE FUELS SUCH AS THE ISO PARAFFIN TYPE HYDROCARBONS-PARTICULARLY WHEN OPERATING UNDER SEVERE CONDITIONS FOR COMBUSTION (I.E. , HIGH AIR FLOW VELOCITY OR LOW COMBUSTION PRESSURE).
4	1	THE ASSUMPTION HAS BEEN MADE IN THIS FUEL EVALUATION THAT THE GREATER THE ALLOWABLE HEAT INPUT RATE FOR A GIVEN VELOCITY, THE GREATER THE DEGREE OF COMBUSTION STABILITY.
4	2	CN THIS BASIS, THE DATA INDICATE THAT ALL THE ADDITIVE MATERIALS TESTED CAUSED AN INCREASE IN STABILITY PERFORMANCE., A FUEL OF RELATIVELY LOW PERFORMANCE SUCH AS TOLUENE BEING BENEFITED TO A GREATER EXTENT THAN A HIGH PERFORMANCE FUEL SUCH AS NORMAL HEPTANE.
4	5	IN GENERAL, ADDITIVE CONCENTRATIONS OF ONE PER CENT BY WEIGHT IN THE SEVERAL PURE HYDROCARBONS WHICH NORMALLY DIFFERED QUITE WIDELY IN PERFORMANCE, PRODUCED UNIFORMLY SUPERIOR COMBUSTION STABILITY CHARACTERISTICS AS MEASURED USING THE PHILLIPS MICROBURNER.
5	0	I. INTRODUCTION
6	1	AT THE REQUEST OF THE NAVY BUREAU OF AERONAUTICS THE JET FUELS GROUP HAS EVALUATED THE EFFECTS OF THE ADDITION OF SMALL AMOUNTS OF DIMETHYLAMINEBORIN =S ON THE COMBUSTION STABILITY PERFORMANCE OF SEVERAL HYDROCARBON FUELS.
7	1	DUE TO THE SMALL QUANTITY OF THIS MATERIAL OBTAINED THE EVALUATION WAS CONDUCTED IN THE PHILLIPS MICROBURNER (MODEL 1A) WHICH IS A SLIGHTLY MODIFIED VERSION OF THE ORIGINAL PHILLIPS MICROBURNER (MODEL 1).
8	0	II. DESCRIPTION OF PHILLIPS MICROBURNER (MODEL 1A)
10	0	III. DESCRIPTION OF TEST APPARATUS
14	0	IV. DESCRIPTION OF TEST FUELS
17	0	V. TEST PROCEDURE
21	0	VI. RESULTS
25	0	VII. DISCUSSION
29	1	PREVIOUS WORK CONDUCTED IN THE PHILLIPS 2 INCH COMBUSTOR (REF. 2) INDICATED THAT SOME ADDITIVES CAUSED A SIGNIFICANT INCREASE IN THE PERFORMANCE OF A LOW RATING FUEL WHILE THESE SAME ADDITIVES DID NOT SUBSTANTIALLY EFFECT THE HIGHER RATING FUELS.
31	3	ALL FOUR ADDITIVES INDICATED THEIR ADDITION TO BE SUBJECT TO THE EFFECT OF DEMINISHING RESULTS UPON FURTHER ADDITION-THAT IS, THEIR EFFECT WAS NOT ESSENTIALLY A BLENDING EFFECT.
33	0	VIII. CONCLUSIONS
34	1	1. THE ADDITION OF DIMETHYLAMINE BORINE IN CONCENTRATIONS OF ONE PER CENT BY WEIGHT TO JET FUEL TYPE HYDROCARBONS RESULTED IN A UNIFORMLY HIGH LEVEL OF COMBUSTION STABILITY PERFORMANCE AS MEASURED BY PHILLIPS MICROBURNER.
35	1	2. THE ADDITION OF RELATIVELY LARGE AMOUNTS OF PROPYLENE OXIDE TO TOLUENE WERE NECESSARY TO PROVIDE SIGNIFICANT IMPROVEMENT IN STABILITY PERFORMANCE AS INDICATED BY INCREASES IN ALLOWABLE HEAT INPUT RATES.
36	1	3. THE ADDITION OF ADDITIVE CONCENTRATIONS (UP TO 1 PER CENT) OF AMYL NITRATE, CUMENE HYDROPEROXIDE, AND DIMETHYLAMINE BORINE ALL RESULTED IN IMPROVED STABILITY PERFORMANCE., THE GREATEST INCREASES WERE SHOWN WHEN BLENDED WITH A FUEL OF POOR PERFORMANCE CHARACTERISTICS-SUCH AS TOLUENE.
37	0	IX. RECOMMENDATIONS
38	1	BASED ON THE EVALUATION OF THE EFFECTS OF ADDITIVES ON THE FLASHBACK LIMITS OF THE ADDITIVE-FUEL BLENDS TESTED IN THE MICROBURNER (MODEL 1A) IT IS RECOMMENDED THAT DIMETHYLAMINE BORINE SHOULD BE FURTHER INVESTIGATED.
38	2	THIS FUTURE WORK SHOULD INCLUDE STUDY OF COMBUSTION STABILITY AND COMBUSTION EFFICIENCY EFFECTS IN THE PHILLIPS 2 INCH COMBUSTOR AND AN INVESTIGATION OF ITS INFLUENCE ON COMBUSTION CLEANLINESS.

Exhibit 11. Automatic Abstract of Document 6 Produced by Combined Method C-T-L

DOCUMENT NUMBER 6 PAGE 1

TEXT	WEIGHT	KEYWORDS
FUEL	28	
COMBUSTION	16	
ADDITIVE	15	
PHILLIPS	14	
STABILITY	14	
BORINE	13	
DIMETHYLAMINE	13	
AIR	11	
CENT	11	
FUELS	11	
HEAT	11	
MICROBURNER	11	
TOLUENE	11	

Exhibit 12. Keywords of Document 6 Ordered by Decreasing Weights Assigned by Key Method K

KEYWORDS DOCUMENT NUMBER 6 PAGE 1

TEXT	WEIGHT
ADDITIVE	15
AIR	11
BORINE	13
CENT	11
COMBUSTION	16
DIMETHYLAMINE	13
FUEL	28
FUELS	11
HEAT	11
MICROBURNER	11
PHILLIPS	14
STABILITY	14
TOLUENE	11

Exhibit 13. Keywords of Document 6 in Alphabetic Order as Produced by Key Method K

PARAGRAPH	SENTENCE	WEIGHT
1	0	1100
2	1	203
2	2	215
2	3	139
3	1	149
3	2	325
4	1	200
4	2	250
4	3	130
4	4	122
4	5	214
5	0	1050
6	1	192
6	2	100
7	1	192
8	0	1040
9	1	104
9	2	74
9	3	50
10	0	1030
11	1	125
11	2	177
11	3	67
12	1	67
13	1	40
14	0	1040
15	1	95
15	2	140
16	1	100
17	0	1000
18	1	90
19	1	38
19	2	20
19	3	- 10
19	4	19
20	1	56
21	0	1040
22	1	95
22	2	121
22	3	67
23	1	67
24	1	117
25	0	1040
26	1	151
26	2	98
27	1	100
27	2	91
28	1	71
28	2	51
28	3	60
28	4	70
29	1	223
29	2	33
29	3	58
29	4	40
30	1	134
30	2	30
30	3	110
31	1	49
31	2	123
31	3	210
32	1	121
32	2	79
33	0	1100
34	1	252
35	1	198
36	1	251
36	2	189
37	0	1090
38	1	294
38	2	241

DOCUMENT NUMBER 6 PAGE 1

Exhibit 14. Sentence Numbers of Document 6 in Textual Order with Weights Assigned by Combined Method C-T-L

DOCUMENT NUMBER 6 PAGE 1

PARAGRAPH	SENTENCE	WEIGHT
1	0	1100
33	0	1100
37	0	1090
5	0	1050
21	0	1040
25	0	1040
14	0	1040
8	0	1040
10	0	1030
17	0	1000
3	2	325
38	1	294
34	1	252
36	1	251
4	2	250
38	2	241
29	1	223
2	2	215
4	5	214
31	3	210
2	1	203
4	1	200
35	1	198
7	1	192
6	1	192
36	2	189
11	2	177
26	1	151
3	1	149
15	2	140
2	3	139
30	1	134
4	3	130
11	1	125
31	2	123
4	4	122
22	2	121
32	1	121
24	1	117
30	3	110
9	1	104
6	2	100
27	1	100
16	1	100
26	2	98
22	1	95
15	1	95
27	2	91
18	1	90
32	2	79
9	2	74
28	1	71
28	4	70
12	1	67
11	3	67
22	3	67
23	1	67
28	3	60
29	3	58
20	1	56
28	2	51
9	3	50
31	1	49
29	4	40
13	1	40
19	1	38
29	2	33
30	2	30
19	2	20
19	4	19
19	3	10

Exhibit 15. Sentence Numbers of Document 6 Ordered by Decreasing Weights Assigned by Combined Method C-T-L

```
     PB   TEXT                 PA1 PA2 P  S  W     CUE WT  KEY WT  TITLE WT                  PAGE     4

START OF DOCUMENT 00C6
    *      EVALUATION                  0  1   1
    *      OF                          0  1   2
    *      THE                         0  1   3
    *      EFFECT                      0  1   4
    *      OF                          0  1   5
    *      CIMETHYLAMINE               0  1   6
    *      BORINE                      0  1   7
    *      ANC                         0  1   8
    *      SEVERAL                     0  1   9
    *      CTHER                       0  1  10
    *      ACCITIVES                   0  1  11
    *      ON                          0  1  12
    *      COMBUSTION                  0  1  13
    *      STABILITY                   0  1  14
    *      CHARACTERISTICS             0  1  15
    *      OF                          0  1  16
    *      VARIOUS                     0  1  17
    *      HYCROCARBON                 0  1  18
    *      TYPE                        0  1  19
    *      FUELS                       0  1  20
    *      IN                          0  1  21
    *      PHILLIPS                    0  1  22
    *      MICROBURNER                 0  1  23
      (    AD87730               )     0  1  24
    CUE WEIGHT        0  KEY WEIGHT      0        TITLE WEIGHT     0   LOCATION WEIGHT    64
AUTHOR
   **      R                  .        0  1   1                        7
   **      L                  .        0  1   2                        7
    *      BRACE                       0  1   3                        7
    CUE WEIGHT        0  KEY WEIGHT      0        TITLE WEIGHT    21   LOCATION WEIGHT    64
SUBHEADING
    *      SUMMARY                     1  0   1      10
    CUE WEIGHT       10  KEY WEIGHT      0        TITLE WEIGHT     0   LOCATION WEIGHT     0
PARAGRAPH

    *      AT                          2  1   1       0
           THE                         2  1   2       0
           RECUEST                     2  1   3
           OF                          2  1   4       0
           THE                         2  1   5       0
    *      NAVY                        2  1   6
    *      BUREAU                      2  1   7
           OF                          2  1   8       0
    *      AERONAUTICS                 2  1   9
    *      PHILLIPS                    2  1  10              14          11
    *      PETROLEUM                   2  1  11
    *      COMPANY                     2  1  12
           UNCERTOOK                   2  1  13
           THE                         2  1  14       0
           EVALUATION                  2  1  15                          11
           OF                          2  1  16       0
           DIMETHYLAMINE               2  1  17              13          11
           BORINE                      2  1  18              13          11
           AS                          2  1  19       0
           AN                          2  1  20       0
           ACCITIVE                    2  1  21              15
           FCR                         2  1  22       0
           IMPROVING                   2  1  23
           THE                         2  1  24       0
           COMBUSTION                  2  1  25              16          11
           CHARACTERISTICS             2  1  26                          11
           OF                          2  1  27       0
           AVIATION                    2  1  28
           GAS                         2  1  29
           TURBINE                     2  1  30
           TYPE                        2  1  31                          11
           FUELS.                      2  1  32
    CUE WEIGHT        0  KEY WEIGHT     71        TITLE WEIGHT    77   LOCATION WEIGHT   126

    *      BECAUSE                     2  2   1      10
           OF                          2  2   2       0
```

Exhibit 16. Portion of Vertical Listing of Document 6 Produced by Combined Method C-K-T-L

3. RESEARCH METHODOLOGY

The various steps of the research methodology are shown in Exhibit 17. The research procedure, stated in its simplest terms, is to

(1) Study the attributes of manual abstracts of documents of a selected portion (Sample Library) of the Experimental Library.

(2) Formulate the definition of abstract suitable for the particular nature of the documents for which success is desired.

(3) Program a machine to produce abstracts and conduct experiments on documents of the Experimental Library.

(4) Test the programs on new documents of the Test Library.

(5) Evaluate the machine abstracts.

3.1 THE RESEARCH PROBLEM

We will begin with a brief summary of the problem of programming a machine to abstract a document. We first assume that an extract of a document (i.e., a selection of certain sentences of the document) can serve as an abstract.* Thus the machine program will be a sentence selection routine which acts on a document stored in machine memory. To create such a program, we will suppose that we have a data base consisting of a collection of documents, each having an abstract prepared by a human. That is to say, an abstractor is given certain instructions concerning desirable features of abstracts and, on the basis of these instructions, prepares an abstract of a document. This gives rise to the problems of the definition of abstracts and the nature of the instructions used to prepare abstracts as well as the related problem of the nature of the

*This hypothesis, i.e., the substitutability of extracts for abstracts is discussed in References 1 and 7. In what follows "abstract" means "extract-type abstract."

document collection. Presuming the successful negotiation of these matters, we consider then the input (i. e., the sentences of the document), and the output (i. e., the selected sentences of the document) of the abstracting program.

Clearly, the machine only can operate on machine-recognizable features of the text. Such features we call characteristics (e. g., occurrence of a certain word, position of sentence in paragraph, number of words in sentence, etc.). A characteristic is said to be relevant if it tends to be associated with selected sentences of the data base (positively relevant) or tends to be associated with unselected sentences (negatively relevant). Using this notion we recast the problem as follows. We (1) find certain relevant characteristics of the text, (2) program the machine to recognize such characteristics, (3) give the machine computational rules to weight sentences according to the presence of these characteristics. These rules are determined from statistical and linguistic considerations concerning the data base, i. e., Sample Library. The final sentence weights are then used as sentence-ranking numbers.

The next step is to examine the output of this procedure and make appropriate adjustments in the weighting schemes. Finally, after making the last adjustments, we test the technique by an evaluation procedure in which machine-produced abstracts are compared with manually-produced abstracts of documents in the Test Library.

In the following sections we discuss the details of the major research areas: The Research Problem (3. 1), Corpus (3. 2), Theory (3. 3), Experiments (3. 4), Evaluation (3. 5). The research steps are diagrammed in Exhibit 17.

3. 2 CORPUS

Documents taken from a particular corpus or body of text have strong similarities among one another and strong dissimilarities with documents taken from a different corpus. This means that one of the first steps in conducting research in automatic abstracting is that of choosing an appropriate corpus. For example, challenges

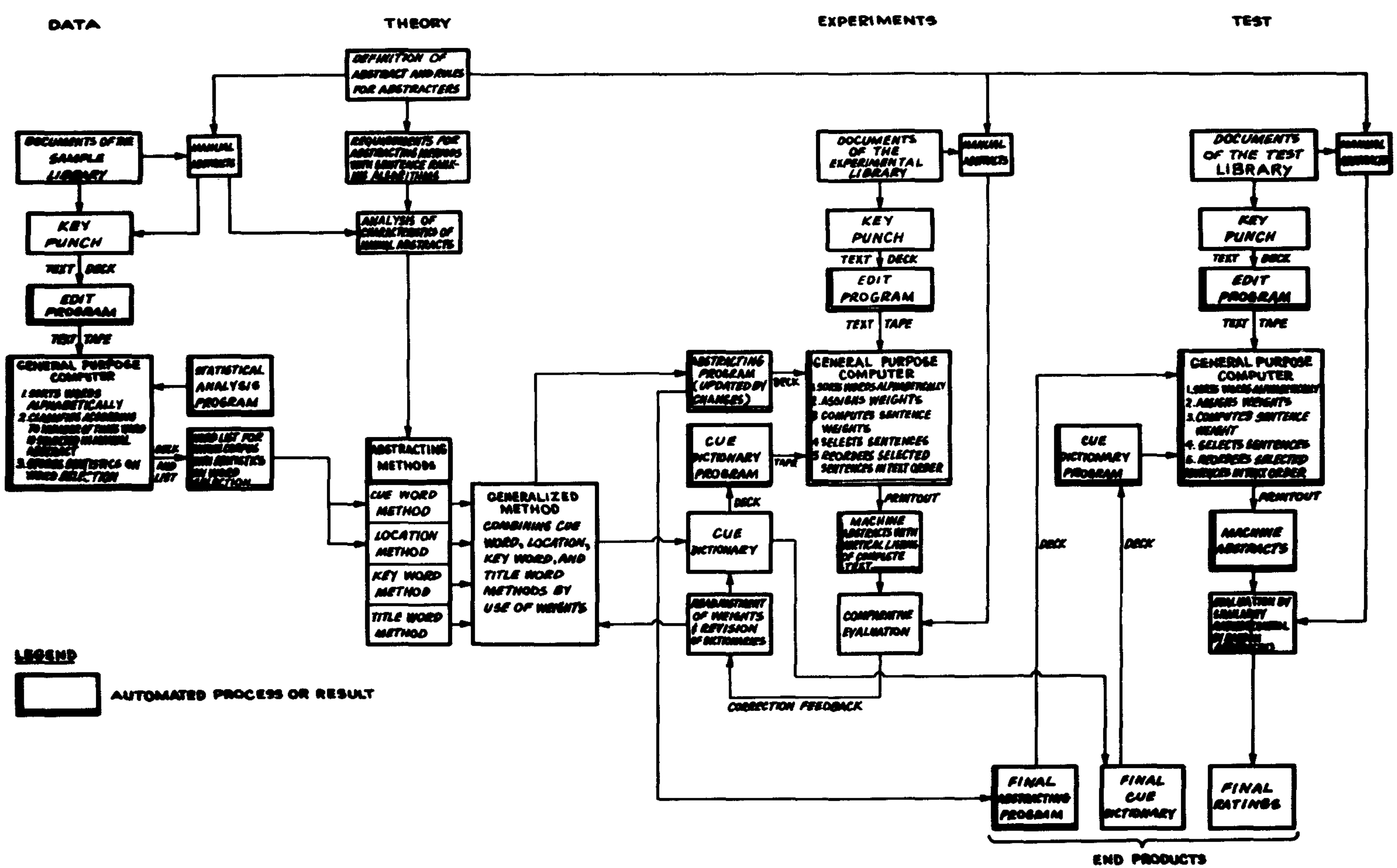


EXHIBIT 17 RESEARCH METHODOLOGY FOR AUTOMATIC ABSTRACTING STUDY

arise due to the subject matter, e.g., education vs. mathematics; to the publishing medium, e.g., newspaper vs. text books; to editors' rules regarding acceptability for publication, e.g., research papers vs. expository works; and to the author's style and compactness of presentation.

3.2.1 Heterogeneous Corpus

The corpus of the previous contract was used in the present study primarily as a basis for research data. The Heterogeneous Corpus contained 200 documents comprising 4 different classes of subject matter: Physical Science, Life Science, Information Science, and Humanities. Because of its heterogeneity and size, this corpus was deemed adequate for some of the statistical data desired, e.g., data on sentence length, data on common words, and positional data. The details of these data are given in Section 3.3.4.

In addition, a special batch TE of 11 documents from the Heterogeneous Corpus was abstracted by the methods developed under the current contract to (1) compare with corresponding abstracts produced under the previous method, and (2) to test the routines designed for a homogeneous corpus (exotic fuel) on a heterogeneous one.

3.2.2 Exotic Fuel Corpus

The corpus used in the present contract consisted of 200 documents dealing with the chemistry of exotic fuels. These 200 documents were obtained from ASTIA as either originals or copies of contractor reports to various government agencies. They were selected by examining over a thousand ASTIA cards and choosing documents whose length did not exceed 4096 words; re-examination showed this threshold to be 3350 words. This condition of length was imposed by the nature of our computer system because of limited storage capacity in the computer memory. Of the 200 selected documents, 85 were unclassified and 115 were confidential.

Since the documents in the Exotic Fuel Corpus were technical reports written by chemists, chemical engineers, and physicists, the style of presentation was that of a typical scientific report, i. e., highly formatted, terse, and containing equations and experimental figures. The lengths of the documents ranged from a minimum of 100 words to a maximum of 3900 words with an average of approximately 2500 words.

For convenience in machine processing, the Exotic Fuel Corpus was divided into several batches which comprised the Experimental Library and the Test Library. The following table presents the details of this batching:

Summary of Batches

Experimental Library (EL)			Test Library (TL)		
Batch	No. of Documents	Classification	Batch	No. of Documents	Classification
A	9	U	TA	20	U
B	20	U	TB	20	U
C	12	U	TC	38	C
D	55	C	TD	22	C
E	4	U			
Total	100			100	

3.2.3 <u>Libraries</u>

As seen in Exhibit 17, which outlines the research methodology for the automatic abstract study, two libraries are necessary. Experimental Library and Test Library are defined as follows: the Experimental Library is a source of documents for the data base and experimentation; the Test Library is a collection of documents reserved for evaluation after experimentation is completed. Documents that are selected from the Experimental Library for detailed study are said to comprise the Sample Library, which serves as the research data base.

The following table summarizes the number of documents and abstracts falling in each of the two libraries for both the Heterogeneous Corpus* and the Exotic Fuel Corpus:

Summary of Libraries

	Documents		Manual Abstracts		Automatic Abstracts	
	Heterogeneous Corpus	Exotic Fuel Corpus	Heterogeneous Corpus	Exotic Fuel Corpus	Heterogeneous Corpus	Exotic Fuel Corpus
Experimental Library	100	100	100	27	20	100
Test Library	100	100	12	40	42	100
Total	200	200	112	67	62	200
	400		179		262	

3.3 THEORY

3.3.1 Definition and Creation of Abstracts

In the course of this research the problem formulation previously derived from a consideration of the nature of abstracts** was reviewed. A reasonable goal of the present study was development

*The meaning of Sample Library and Experimental Library used in the present study is not that of Reference 7.

**See References 1 and 7.

of a method of machine-producing indicative-type abstracts for use in the screening of documents. Because of the special nature of the Exotic Fuel Corpus it was decided that such a function of abstracts could be achieved by specifying the information content of abstracts of technical papers. With this goal in mind, a semiformal explication of the notion of abstract was developed. This comprises a sequence of four definitions: (1) Eligible Sentence*, (2) Nonredundant Sentences, (3) Coherence, and (4) Abstract.

Definition 1. Eligible Sentence. A sentence is called eligible if it contains information of at least one of the following six types:

S Subject Matter. Information indicating the general subject area that is the author's principal concern; i. e., what?

P Purpose. Information indicating whether the author's principal intent is to offer original research findings, to survey or evaluate the work performed by others, to present a speculative or theoretical discussion, or to serve some other main purpose; i. e., why?

M Methods. Information indicating the methods used in conducting the research. Depending on the type of research, such statements may refer to experimental procedures, mathematical techniques, or other methods of scientific investigation; i. e., how?

C Conclusions or Findings. Information indicating the results obtained in the research or the findings of the author.

G Generalizations or Implications. Information indicating the significance of the research and its bearing on broader technical problems or theory.

R Recommendations or Suggestions. Information indicating recommended courses of action or suggested areas of future work.

---

*The neutral word "eligible" is used rather than the word "significant."

Definition 2. Nonredundant Sentence. A sentence or group of sentences will be called nonredundant if it has none of the following properties:

(1) It is an exact replica of another sentence or a smaller group of sentences.

(2) It is a paraphrase of another sentence or a smaller group of sentences.

(3) It expresses the same content as another sentence or a smaller group of sentences.

Definition 3. Coherence. A sequence of sentences is said to be coherent if it has the following properties:

(1) All crucial antecedents and referents are present.

(2) No semantic discontinuities are present.

(3) The sequence of ideas progresses logically.

Definition 4. Abstract. A sequence of sentences of a document selected in text order is said to be an abstract of the document if it has the following properties:

(1) Property of Content: It contains only eligible sentences.

(2) Property of Length: It contains nonredundant sentences.

(3) Property of Form: It is coherent.

A set of Instructions for Abstractors (Exhibit 18) was then developed in order to create abstracts satisfying the above definition. Abstracts generated in accordance with these instructions then serve as the target abstracts required by the research methodology. Because of the requirements of a mechanized technique, it was stipulated that the target abstracts contain a fixed percent of the sentences of the original document. We chose 25 percent, but believe that is not optimum.

During the preparation of these abstracts a document analysis sheet was also prepared. This gives an itemization of the sentences with (1) notations as to the occurrence of cointensional sentences, (2) classification of sentences into the six information categories, and (3) qualitative judgments as to the abstract-worthiness of sentences. The selected sentences are then checked on the analysis sheet in the Column H ("H" for "human"). A Column R is also provided to record a random selection of sentences. The randomly generated extract, so defined, is then used as a second control in the evaluation procedure. Exhibit 19 gives the completed analysis sheet for Exotic Fuel Document 6 (Exhibit 3). The corresponding Target Abstract is given in Exhibit 20 and a Random Extract in Exhibit 21.

A major part of the abstracting effort was devoted to locating duplications of information throughout the document. A typical abstract in the corpus examined presented a summary at the beginning, a section of orientation, a section of methods, a discussion section, and frequently conclusions which were a rephrasing of the initial summary. Hence the same information frequently appeared two or three places in a document. Occasionally, however, the opposite situation prevailed, wherein nearly the entire document comprised conclusions of the research. When a hierarchy cannot be established, either of significance or of generality, then a 25 percent abstract in the sense of a sampling of content cannot be logically composed. In this case a description is the appropriate representation. If the document contains one it is rarely 25 percent of the sentences; if it does not, the present extracting method will not produce a coherent abstract. Also, in order to specify the full 25 percent of the sentences, occasionally it was necessary to select several less condensed sentences in place of the single more succinct, more "meaty" one which was their equivalent. Although a 25 percent selection seemed workable for most of the documents, in some instances it was not appropriate, most frequently by exceeding optimum length.

Step 1.	NUMBER:	Assign numerical designations to the sentences of the document.
Step 2.	ORIENT:	Read the definitions of eligibility, non-redundancy, coherence, and abstract.
Step 3.	FAMILIARIZE:	Read the article quickly to gain familiarity.
Step 4.	CLASSIFY:	Reread the article slowly, classifying successive sentences if eligible. Record the class on the document analysis sheet. If a sentence is ineligible or obviously not abstract-worthy, mark nothing.
Step 5.	LOCATE DEPENDENCIES:	Note with an arrow on the analysis sheet if an antecedent sentence would have to be selected also in order to preserve the meaning.
Step 6.	LOCATE REDUNDANCIES:	Reread the article to locate redundancies and note them in as great detail as necessary on the analysis sheet.
Step 7.	SELECT:	Select 25 percent of the number of sentences in the document by applying the principle of coherence and a comparative notion of sentence importance. Avoid redundancy.
Step 8.	REVIEW:	Reread the finished abstract to check its coherence, conformity to prescribed length, and freedom from redundancy.

Exhibit 18. Instructions for Abstractors

NOTES	PARAGRAPH	SENTENCE	CLASS	H	CTL	C	T	L	K	CTLK	R
SUMMARY	1	0									
≈ 6-1	2	1	S/P	X	X		X	X	X	X	
> 6-1 + 6-2 > 7-1		2	S	X	X		X	X		X	
		3	S	X				X			
≈ 16-1	3	1	S/M	X				X			–
≈ 4-3 > 29-1 "such materials"		2	P	X	X	X	X	X	X	X	X
= 26-2	4	1	M		X	X		X	X	X	
< 30-3 ≈ 36-1,2 "on this basis"		2	C		X	X		X	X	X	
≈ 3-2 ≈ 29-1 "these data"		3	G					X			X
< 4-5 ≈ 31-2 < 34-1 ≈ 35-1		4	C/G	X				X	X	X	
> 4-4 ≈ 34-1		5	C	X	X	X	X	X	X	X	
INTRODUCTION	5	0									
≈ 2-1	6	1	S		X	X	X			X	
< 2-2		2									X
< 2-2	7	1	M		X		X			X	
DESCRIPTION OF MICROBURNER	8	0									
	9	1									
		2									
		3									X
DESCRIPTION OF TEST APPARATUS	10	0									
	11	1	M								X
		2	M			X					
"the new system"		3	M								
	12	1									
"these modifications"	13	1									X
DESCRIPTION OF TEST FUELS	14	0									
	15	1									
"these fuels"		2	G	X		X					X
≈ 2-3 + 3-1 > 25-1	16	1	M				X				
TEST PROCEDURE	17	0									
	18	1	M*								
	19	1									
		2									X
		3									
		4									
	20	1									X

* Significant content removed by pre-editing.

NOTES	PARAGRAPH	SENTENCE	CLASS	H	CTL	C	T	L	K	CTLK	R
RESULTS	21	0									
	22	1	C*								X
		2	M			X					
		3	M						X		
	23	1	M								X
	24	1	M	X		X					
DISCUSSION	25	0									
	26	1	M				X		X		
= 4-1		2	M/C	X							
	27	1	M								
		2	C/G								
< 16-1	28	1	M								
		2	C*								
		3	C								
"thus"		4	M								X
< 3-2 ≈ 4-3	29	1	P		X	X	X			X	
"therefore"		2	M						X		X
		3	M/P								X
		4	C*								
	30	1	P/M				X				
		2	C								
> 4-2 ≈ 36-1,2		3	C/G			X					
"these curves"	31	1	M/P						X		
≈ 4-4 = 34-1 + 35-1		2	C			X					
		3	C	X	X	X	X				X
	32	1	C	X			X		X		
		2	C								
CONCLUSIONS	33	0									
> 4-4 ≈ 4-5 ≈ 4-4,5	34	1	C		X	X	X	X	X	X	
≈ 4-4 = 31-2	35	1	C		X	X		X		X	
≈ 4-2,3	36	1	C	X	X	X	X	X	X	X	
≈ 30-3		2	C	X		X		X			X
RECOMMENDATIONS	37	0									
	38	1	R	X	X	X	X	X	X	X	
		2	R	X	X	X	X	X	X	X	

* The sentence refers to parts of the document removed by pre-editing.

Exhibit 19. Analysis of Document 6 and Comparison of Human and Machine Abstracts

Notes on Exhibit 19

1. Abstracts are designated according to the method of selection:

H	Human (elsewhere referred to as the "Target" abstract)
CTL	Cue-Title-Location combined method, chosen as the preferred method
C	Cue method
T	Title method
L	Location method
K	Key method
CTLK	Cue-Title-Location-Key combined method
R	Random selection

2. All headings (designated by a "0" sentence number) are output in all machine abstracts, although not checked in the table. They have been inserted in the Target Abstract and Random Extract (Exhibits 20 and 21) for purposes of comparison with the machine abstracts.

3. The class designations of sentences are:

S	Subject
P	Purpose
M	Method
C	Conclusion
G	Generalization
R	Recommendation

4. The following symbols are used to indicate cointensional sentences. Sentences are identified by the paragraph and sentence number assigned by the computer.

$=$	is equivalent to
$\approx$	is approximately equivalent to
$>$	contains more information than
$<$	contains less information than
$\}$	the two sentences taken together

5. An arrow indicates that for preservation of meaning an antecedent sentence is required. Note is made, when possible, of the words indicating this dependence.

6. Sentences marked with an asterisk are no longer meaningful because they contain or refer to essential symbols, equations, tables, figures, and graphs which have been deleted in the pre-editing step.

Exhibit 19. Analysis of Document 6 and Comparison of Human and Machine Abstracts (Continued)

PAR SENT DOCUMENT NUMBER 6 1

ABSTRACT BASED ON HUMAN SELECTION

EVALUATION OF THE EFFECT OF DIMETHYLAMINE BORINE AND SEVERAL OTHER ADDITIVES ON COMBUSTION STABILITY CHARACTERISTICS OF VARIOUS HYDROCARBON TYPE FUELS IN PHILLIPS MICROBURNER (AD87730)

R. L. BRACE

PAR	SENT	
1	0	SUMMARY
2	1	AT THE REQUEST OF THE NAVY BUREAU OF AERONAUTICS, PHILLIPS PETROLEUM COMPANY UNDERTOOK THE EVALUATION OF DIMETHYLAMINE BORINE AS AN ADDITIVE FOR IMPROVING THE COMBUSTION CHARACTERISTICS OF AVIATION GAS TURBINE TYPE FUELS.
2	2	BECAUSE OF THE SAMLL AMOUNT (100 GRAMS) OF DIMETHYLAMINE BORINE RECEIVED FROM CALLERY CHEMICAL COMPANY, THIS EVALUATION HAS BEEN LIMITED TO THE MEASUREMENT OF ITS EFFECT ON THE FLASH-BACK CHARACTERISTICS OF THREE PURE HYDROCARBONS (TOLUENE, NORMAL HEPTANE AND BENZENE) IN THE PHILLIPS MICROBURNER.
2	3	DIMETHYLAMINE BORINE CONCENTRATIONS OF FROM 0.1 to 1.0 PER CENT BY WEIGHT WERE EVALUATED.
3	1	FOR COMPARATIVE PURPOSES TWO COMMON IGNITION ADDITIVES (AMYL NITRATE AND CUMENE HYDROPEROXIDE) WERE ALSO EVALUATED DURING THIS STUDY, AS WELL AS CONCENTRATIONS UP TO 20 PER CENT BY WEIGHT OF PROPYLENE OXIDE - A RELATIVELY HIGH FLAME VELOCITY FUEL.
3	2	PREVIOUS STUDIES IN PHILLIPS 2 INCH TURBOJET ENGINE TYPE COMBUSTOR HAD INDICATED THAT SUCH MATERIALS COULD SUBSTANTIALLY INCREASE THE MAXIMUM RATE OF HEAT RELEASE ATTAINABLE, ESPECIALLY WITH LOW PERFORMANCE FUELS SUCH AS THE ISO PARAFFIN TYPE HYDROCARBONS - PARTICULARLY WHEN OPERATING UNDER SEVERE CONDITIONS FOR COMBUSTION (I. E., HIGH AIR FLOW VELOCITY OR LOW COMBUSTION PRESSURE).
4	4	WITH RESPECT TO THE DIMETHYLAMINE BORINE, ITS EFFECT AS A FUEL ADDITIVE WAS NOTEWORTHY; 0.1 WEIGHT PER CENT IN TOLUENE BEING EQUIVALENT TO 20 PER CENT BY WEIGHT OF ADDED PROPYLENE OXIDE.
4	5	IN GENERAL, ADDITIVE CONCENTRATIONS OF ONE PER CENT BY WEIGHT IN THE SEVERAL PURE HYDROCARBONS WHICH NORMALLY DIFFERED QUITE WIDELY IN PERFORMANCE, PRODUCED UNIFORMLY SUPERIOR COMBUSTION STABILITY CHARACTERISTICS AS MEASURED USING THE PHILLIPS MICROBURNER.
5	0	I. INTRODUCTION
8	0	II. DESCRIPTION OF PHILLIPS MICROBURNER (MODEL 1A)
10	0	III. DESCRIPTION OF TEST APPARATUS
14	0	IV. DESCRIPTION OF TEST FUELS
15	2	THESE FUELS REPRESENT VARIATIONS IN CHEMICAL STRUCTURE WHICH WILL IN TURN PROVIDE INDICES OF BOTH GOOD AND POOR COMBUSTION STABILITY PERFORMANCE.
17	0	V. TEST PROCEDURE
21	0	VI. RESULTS
24	1	THE REGION OF STABLE OPERATION IS DEFINED AS THE STATE OF FLASH BACK— THE CONDITIONS OF COMBUSTION WHERE THE FLAME WOULD BECOME ANCHORED TO A FLAME HOLDER - AS IN STABLE GAS TURBINE OR RAM JET COMBUSTOR OPERATION - IF THE FLAME HOLDER WERE PROVIDED IN THE BURNER TUBE.
25	0	VII DISCUSSION
26	2	THE ASSUMPTION IS MADE THAT THE GREATER THE ALLOWABLE BEAT INPUT RATE AT A GIVEN VELOCITY, THE GREATER THE DEGREE OF STABILITY.
31	3	ALL FOUR ADDITIVES INDICATED THEIR ADDITION TO BE SUBJECT TO THE EFFECT OF DEMINISHING* RESULTS UPON FURTHER ADDITION - THAT IS, THEIR EFFECT WAS NOT ESSENTIALLY A BLENDING EFFECT.
32	1	MENTION SHOULD BE MADE OF THE FACT THAT DURING THE COMBUSTION OF THE DIMETHYLAMINE BORINE-HYDROCARBON FUEL BLENDS NO NOTICEABLE ODORS OR SMOKE WERE OBSERVED.
33	0	VIII. CONCLUSIONS
36	1	3. THE ADDITION OF ADDITIVE CONCENTRATIONS (UP TO 1 PER CENT) OF AMYL NITRATE, CUMENE HYDROPERIOXIDE, AND DIMETHYLAMINE BORINE ALL RESULTED IN IMPROVED STABILITY PERFORMANCE; THE GREATEST INCREASES WERE SHOWN WHEN BLENDED WITH A FUEL OF POOR PERFORMANCE CHARACTERISTICS - SUCH AS TOLUENE.
36	2	BENEFICIAL EFFECTS WERE APPRECIABLY LESS WHEN BLENDED WITH A FUEL OF GOOD PERFORMANCE CHARACTERISTICS - SUCH AS N-HEPTANE.
37	0	IX. RECOMMENDATIONS
38	1	BASED ON THE EVALUATION OF THE EFFECTS OF ADDITIVES ON THE FLASHBACK LIMITS OF THE ADDITIVE-FUEL BLENDS TESTED IN THE MICROBURNER (MODEL 1A) IT IS RECOMMENDED THAT DIMETHYLAMINE BORINE SHOULD BE FURTHER INVESTIGATED.
38	2	THIS FUTURE WORK SHOULD INCLUDE STUDY OF COMBUSTION STABILITY AND COMBUSTION EFFICIENCY EFFECTS IN THE PHILLIPS 2 INCH COMBUSTOR AND AN INVESTIGATION OF ITS INFLUENCE ON COMBUSTION CLEANLINESS.

*Sic

Exhibit 20. Target Abstract

PAR	SENT	DOCUMENT NUMBER 6 1
		EXTRACT BASED ON RANDOM SELECTION EVALUATION OF THE EFFECT OF DIMETHYLAMINE BORINE AND SEVERAL OTHER ADDITIVES ON COMBUSTION STABILITY CHARACTERISTICS OF VARIOUS HYDROCARBON TYPE FUELS IN PHILLIPS MICROBURNER (AD87730) R. L. BRACE
1	0	SUMMARY
3	2	PREVIOUS STUDIES IN PHILLIPS 2 INCH TURBOJET ENGINE TYPE COMBUSTOR HAD INDICATED THAT SUCH MATERIALS COULD SUBSTANTIALLY INCREASE THE MAXIMUM RATE OF HEAT RELEASE ATTAINABLE, ESPECIALLY WITH LOW PERFORMANCE FUELS SUCH AS THE ISO PARAFFIN TYPE HYDROCARBONS - PARTICULARLY WHEN OPERATING UNDER SEVERE CONDITIONS FOR COMBUSTION (I. E., HIGH AIR FLOW VELOCITY OR LOW COMBUSTION PRESSURE).
4	3	THESE DATA ARE IN AGREEMENT WITH PREVIOUS ADDITIVE STUDIES BY PHILLIPS.
5	0	I. INTRODUCTION
6	2	THE DIMETHYLAMINE BORINE WAS SUPPLIED TO PHILLIPS BY THE CALLERY CHEMICAL COMPANY.
8	0	II. DESCRIPTION OF PHILLIPS MICROBURNER (MODEL 1A)
9	3	THE DETAILS OF THE MODEL 1A ARE SHOWN IN FIGURE 1.
10	0	III. DESCRIPTION OF TEST APPARATUS
11	2	IN THE PRESENT EVALUATION IT WAS NOT NECESSARY TO CONSIDER THE EFFECT OF CORROSION, CONSEQUENTLY A CONTINUOUS FLOW SYSTEM PROVIDING GREATER FLEXIBILITY AND EASIER HANDLING WAS INCORPORATED WHICH REQUIRES ONLY SLIGHTLY MORE FUEL PER TEST THAN THE ORIGINAL.
13	1	THE DETAILS OF THESE MODIFICATIONS AND OF THE TEST APPARATUS ARE SHOWN IN SCHEMATIC IN FIGURE 2.
14	0	IV. DESCRIPTION OF TEST FUELS
15	2	THESE FUELS REPRESENT VARIATIONS IN CHEMICAL STRUCTURE WHICH WILL IN TURN PROVIDE INDICES OF BOTH GOOD AND POOR COMBUSTION STABILITY PERFORMANCE.
17	0	V. TEST PROCEDURE
19	2	IGNITION OF THE THEN FUEL-RICH MIXTURE WAS ACCOMPLISHED BY APPLYING A LIGHTED, PORTABLE PROPANE TORCH TO THE TOP OF THE BURNER TUBE.
20	1	AFTER CHECKING THE POINT AT LEAST ONCE MORE THE AIR FLOW WAS INCREASED ANOTHER INCREMENT AND THE PROCESS REPEATED.
21	0	VI. RESULTS
22	1	THE RESULTS OF THE EVALUATION OF FLASHBACK LIMITS OF THE FUELS AND FUEL-ADDITIVE BLENDS ARE SUMMARIZED IN TABLE 2 AND SHOWN GRAPHICALLY IN FIGURE 3 THROUGH 8.
23	1	THE REFERENCE VELOCITY IS DETERMINED BY THE AIR FLOW CONDITIONS AT ENTRY TO THE BURNER TUBE - NEGLECTING THE MASS OF THE FUEL PARTICLES.
25	0	VII. DISCUSSION
28	4	THIS EVALUATION SERVED PRIMARILY AS A REFERENCE PLANE OF PERFORMANCE IMPROVEMENT.
29	2	THEREFORE, IN ADDITION TO THE EVALUATION OF THE FUEL BLENDS CONTAINING 0.1, 0.5, AND 1 PER CENT DIMETHYLAMINE BORINE IN TOLUENE (SHOWN IN FIGURE 5) THE SAME ADDITIVE WAS TESTED IN N-HEPTANE (SHOWN IN FIGURE 6).
29	3	VARIFICATION* OF THE PRIOR RESULTS WAS FURTHER ESTABLISHED BY TESTING TWO OF THE ADDITIVES PREVIOUSLY EVALUATED (REF 2.)
31	3	ALL FOUR ADDITIVES INDICATED THEIR ADDITION TO BE SUBJECT TO THE EFFECT OF DEMINISHING* RESULTS UPON FURTHER ADDITION - THAT IS, THEIR EFFECT WAS NOT ESSENTIALLY A BLENDING EFFECT.
33	0	VIII. CONCLUSIONS
36	2	BENEFICIAL AFFECTS WERE APPRECIABLY LESS WHEN BLENDED WITH A FUEL OF GOOD PERFORMANCE CHARACTERISTICS - SUCH AS N-HEPTANE.
37	0	IX. RECOMMENDATIONS

*Sic

Exhibit 21. Random Extract

An attempt was made to classify qualifying sentences as to degree of "abstract-worthiness" (a concept analogous to that of the machine weightings), but in practice it did not prove a sufficient method for sentence selection. It is often impossible to tell if a sentence, seen only in the context of an abstract, reports well-known fact, the result of previous experiment, or if it is a conclusion of the present study. Furthermore, a sentence frequently depends on a previous one for its meaning (and there may be a series of such dependencies) requiring selection of both, even though the antecedent sentence may not qualify in terms of content. Certain words or phrases frequently indicate this situation, such as "this," "therefore," "since," etc., and a study of their occurrence might be useful in refining the machine method. The more an author uses such words for reasons of logical structuring and stylistic fluidity, the more difficult it is to remove sentences from context without destroying their function in the document.

In short, it is found that in composing as coherent and meaningful an abstract as possible, requirements of antecedents, absence of elements eliminated by pre-editing, suppression of redundancy, and considerations related to the length quota often take precedence over a sentence-by-sentence rating of "abstract-worthiness." It is in precisely these aspects that the human extracts differ consistently from the machine-produced ones.

3.3.2 Guiding Principles

A set of principles was next devised to guide the development of automatic abstracting methods. Among the several principles, we stress one that seems dominant:

Principle 1. Insure that the automatic method detects and uses all abstracting clues (e.g., of meaning, significance, organization, etc.) provided by the author, editor, and printer.

This principle focuses on capturing automatically as many clues as possible that were, either consciously or unconsciously, provided by the creators of the document. For example, the skilled author selects an appropriate title, organizes his thoughts in distinct sections with appropriate subtitles, condenses information in the captions of graphs and tables, and uses footnotes and references in revealing ways.

It is instructive to regard the problem of automatic abstracting in the light of several other principles:

Principle 2.	Use criteria of selection, i.e., a system of rewards for desired sentences.
Principle 3.	Use criteria of rejection, i.e., a system of penalties for undesired sentences.
Principle 4.	Use a system of thresholds, both for acceptance and rejection, that allows adjustments by parameterization.
Principle 5.	Use a method which is a function of several distinct factors, such as statistical, semantic, syntactic, locational, etc.

It thus appears that an abstracting system based on assigning numerical weights (weights above a threshold for positively relevant characteristics, at the threshold for irrelevant characteristics, below the threshold for negatively relevant characteristics) to machine-recognizable sentence characteristics, will satisfy these principles. For computational simplicity, addition of the weights is used to arrive at the final ranking numbers.

In the next section we describe four basic methods which rely upon the five principles stated above. It should be noted, however, that in regard to Principle 1 we developed methods utilizing clues in the title and subtitles but were unable, because of limited time and scope, to program in the operating system clues known to exist in captions of tables, footnotes, and references.

3.3.3 The Four Basic Methods

The automatic abstracting system developed under this contract is founded on four basic methods: Key, Cue, Title, Location. Their origin will now be described.

First, the clues of the document may come from two structural sources.

(1) Clues in the skeleton of the document, e.g., titles, headings, format.

(2) Clues in the body of the document, e.g., the text.

Second, the characteristics of the corpus may be considered from two linguistic points of view.

(1) General characteristics of the entire corpus; e.g., certain function words.

(2) Specific characteristics of the individual documents; e.g., high frequency content words.

These two sources of clues and two types of linguistic properties yield four opportunities to create distinct basic methods of automatic abstracting defined simply by the class of clues they rely upon. They are displayed below:

Rationale of the Four Basic Methods

Type of Linguistic Property	Sources of Structural Clues	
	Body of Document (Text)	Skeleton of Document (Title, Headings, Format)
General Characteristics of Corpus	Cue Method	Location Method
Specific Characteristics of Document	Key Method	Title Method

When this classification is applied to words (considered as clues) it yields four distinct word lists. Of these, we distinguish between two different types of word lists.

First: A dictionary is a word list plus numerical tags which forms a fixed input to the automatic abstracting system; thus a dictionary is independent of the words in the document being abstracted.

Second: A glossary is a word list plus numerical tags which forms a variable input to the automatic abstracting system; thus a glossary is dependent on the words of the document being abstracted.

This refinement and standardization of terminology has provided a convenient breakdown of the word lists that correspond to each of the four basic methods. The table below shows this relationship:

Rationale of the Four Word Lists

	Body of Document	Skeleton of Document
Function Words	Cue Dictionary (Bonus, Stigma, and Null Dictionaries)	Heading Dictionary
Content Words	Key Glossary	Title Glossary

The method of generating the Key and Title Glossaries given the Cue Dictionary has been described in Section 2.

Exhibit 22 presents an inventory of the sentence characteristics considered in the course of the research. The rules governing their use in the abstracting programs developed have been given in Section 2.

Sentence Characteristic	Method	Relevance Type	Verification of Relevance Type	Weight	Derivation of Weight
Contains Bonus Word of Cue Dictionary	Cue	Positive	Statistical & Linguistic	+10	Linguistic
Contains Null Word of Cue Dictionary	Cue	Irrelevant	Statistical & Linguistic	0	Linguistic
Contains Stigma Word of Cue Dictionary	Cue	Negative	Statistical & Linguistic	-10	Linguistic
Contains a Title word of the Title Glossary	Title	Positive	Statistical	11	Linguistic & Mathematical
Contains a Heading word of the Title Glossary	Title	Positive	Statistical	7	Linguistic & Mathematical
Contains a word of the Key Glossary	Key	Positive	Linguistic	Frequency of word	Linguistic
Occurs under a heading containing a word of the Heading Dictionary	Location	Positive	Statistical & Linguistic	See Heading Dictionary	Statistical & Linguistic
Occurs in the first paragraph of a section	Location	Positive	Statistical & Linguistic	18	Statistical & Linguistic
Occurs in the last paragraph of a section	Location	Positive	Statistical & Linguistic	18	Statistical & Linguistic
Is the first sentence of a paragraph	Location	Positive	Statistical & Linguistic	9	Statistical & Linguistic
Is the last sentence of a paragraph	Location	Positive	Statistical & Linguistic	9	Statistical & Linguistic

Exhibit 22 - Inventory of Sentence Characteristics

3.3.4 Characteristics, Dictionaries, and Weights

(1) Cue Dictionary. The previous research used two distinct Cue Dictionaries: (1) a dictionary based on purely statistical considerations, (2) a dictionary based on purely linguistic considerations.* In the present research it was decided to base the new Cue Dictionary on a combination of statistical and linguistic properties by taking the following steps.

Step 1. Selection of Candidates for Cue Words. The Sample Library of the Heterogeneous Corpus used in the previous research consisted of 16,386 different words. The output data of the Concordance Program* gave the following statistics for each word: (1) frequency in corpus; (2) number of documents in which word occurred (dispersion); (3) selection ratio (ratio of occurrences in abstractor-selected sentences to frequency in corpus). Tabulating equipment was used to separate the words having dispersion less than 5 from those having dispersion 5 or greater. The latter class consisted of 3,314 words which were deemed to be the source of function words of the language and hence candidates for Cue words.

Step 2. Classification of Candidates. Two statistical thresholds were established above and below the mean selection ratio for all words. The following classes were then defined and listed by means of a computer program:

Null candidates:	dispersion greater than 30 and selection ratio between thresholds yielded 282 words
Bonus candidates:	selection ratio above upper threshold yielded 986 words
Stigma candidates:	selection ratio below lower threshold yielded 1,177 words
Residue:	dispersion less than 30 and selection ratio between thresholds yielded 869 words

---

*See Reference 7.

Step 3. Review and Compile Dictionary, Version 1. The listings of Step 2 were reviewed, and reclassified when necessary. Two conditions guided this effort: (1) assignments must not be counterintuitive; (2) the computer program allows only 1000 words in the Cue Dictionary. It was found that most of the counterintuitive assignments occurred when the word frequency was low, giving an unreliable selection ratio. For high frequency words, it was found that intuition usually bore out the statistical data. The second condition was found to be easily met by considering the probability of occurrence of a word. That is to say, it is possible to adhere to the Cue Dictionary size limitation by casting out certain low frequency words without fear of loss of effectiveness. It turned out that most of the Stigma candidates were reclassified as residue by this criterion. A dictionary was compiled of the Null, Bonus, and Stigma words remaining. Also, a new class of indeterminate words was formed and listed for further study. The breakdown was: 324 Null words, 568 Bonus words, 21 Stigma words, 314 indeterminate words.

Step 4. Experimentation and Analysis. The Cue Dictionary, Version 1 was used in 3 abstracting experiments (1-3)* on a sample of exotic fuel documents. The Key word and vertical listing output were then studied for occurrences of Null, Bonus-Stigma words and candidates for additional Cue Dictionary entries. Obvious errors were noted and corrected, resulting in the Cue Dictionary, Version 2.

This first examination of the effect of the Dictionary led to two decisions: (1) to formulate linguistic guidelines for further classification oriented more specifically to the Exotic Fuel Corpus; (2) to create a statistical data base (as done on the Heterogeneous Corpus) from 20 documents of the Exotic Fuel Corpus, giving selection ratio and frequency data.

---

*See Section 3.4. The operating system at this stage of the research did not have the capability of handling negative dictionary weights. Thus, the Stigma words were in the Cue Dictionary but not used.

Step 5. Linguistic Reorientation. The linguistic analysis of the experimental results led to the following descriptions of word classifications.

Null

ordinals	pronouns
cardinals	adjectives
verb "to be"	verbal auxiliaries
prepositions	articles
verbs of state or process	coordinating conjunctions

Bonus

comparatives	relative interrogatives
superlatives	causality terms
adverbs of conclusion	important conditions or processes
value terms	

Stigma

belittling expressions	plurals of explicatory expressions
references elsewhere	hedging expressions
insignificant-detail expressions	

Residue

positives
technical terms
archaic terms

Version 2 was now reviewed according to these guidelines and Version 3 produced. The members of the indeterminate class were given a final adjudication. The breakdown of this Dictionary is 164 Null words, 789 Bonus words, 47 Stigma words. Version 3 of the Cue Dictionary was used in Experiments 4-11. *

Step 6. Statistical Reorientation. The statistics for the exotic fuel sample (Batch B) were examined. If a word had frequency greater than or equal to 25 it was assigned to the Bonus class when the selection ratio was high, to the Stigma class if the selection ratio was low. The medium selection-ratio words with

---

*See Section 3. 4. The reprogramming was completed for the use of Stigma words in these experiments.

exceptionally high frequency were assigned to the Null class. The results of this mechanical step were reviewed for violations of linguistic plausibility. Such violations were rare. That is to say, the statistics confirmed to a large degree the previous dictionary entries. Finally, the machine abstracts of Experiment 11 were examined. It was observed that some words were improperly called Key words, in particular, units of measurement. It was then discovered that, by gathering all word occurrences of units of measurement into a single class, the ratio of selection of this class was very low. Thus units of measurement were assigned to the Stigma class. Deletions were made to allow room for the new entries according to the criterion of low probability of occurrence in the Exotic Fuel Corpus. The final version (Version 4) was then formed to incorporate these changes. The breakdown into final dictionaries is: 139 Null words, 783 Bonus words, 73 Stigma words.

The Cue weights are summarized in Exhibit 22, and a section of the Cue Dictionary is shown in Exhibit 4.

(2) Heading Dictionary and Ordinal Weights. The motivation behind the use of location factors stems from two considerations: The first is that we can postulate that if a "topic sentence" exists in a paragraph it will tend to occur early or late in the paragraph. The second is that, in a technical report, a sentence that occurs under certain headings (e. g. "Purpose", "Conclusions") has a strong chance of being suitable for extraction. Consequently a sentence should receive a reward for its position in a paragraph and for its occurrence under certain headings. The plan was to list certain location characteristics of sentences and then to test whether such characteristics were indeed relevant factors. Finally, weights were assigned to the characteristics. The two types of location factors mentioned above gave rise to the Heading Dictionary and the Ordinal weights. The compilation of the Heading Dictionary will be discussed first. The work proceeded by utilizing both linguistic and statistical considerations (as in the Cue Dictionary compilation).

Step 1. Data Collection and Preliminary Classification. Each heading occurring in 100 exotic fuel articles was listed separately on a card together with the frequency of occurrence of the heading. These headings were then classified by a scheme developed in terms of the defined information content of abstracts. The classification was: S: subject matter headings; P, purpose headings: M, method headings; C, conclusion headings; G, generalization headings; and R, recommendation headings. That is to say, on the basis of the headings classified under an S-classed heading we expect to find general subject matter sentences, under a P-classed heading we expect to find sentences concerning the purpose of the article, etc. Another class F was defined comprising headings of a purely functional nature under which we expect to find summarizing sentences (e.g., "Abstract", "Summary"). The remaining cards were classified as: N, no use in abstract (e.g. "Appendix"); I, indeterminate (classification is possible, but the correct assignment is doubtful); U, unclassified (no classification applies).

Step 2. Organization of Word Data. An alphabetic listing was was made of all words that occurred in the headings with the exception of prepositions, articles, and highly specific words.* The words were given with frequency of occurrence, frequency in combination, total frequency, and distribution data of the classification of step 1; 156 words were thus obtained.

Step 3. Statistical Augumentation. Next, 20 documents of the Exotic Fuel Corpus (Batch B) were selected for detailed study. An outline of each document was prepared showing the number of sentences occurring under each heading. The target abstracts of these documents were then examined for the number of sentences selected under each heading. It was found that 31 heading words of the previous listing occurred in this document sample. For each of

*Content words were included if they had the slightest generality (e.g., "facilities", "materials"). That is to say, caution was exercised in preserving the peculiarities of the corpus.

these words the selection ratio was computed, i.e., the ratio of the number of abstractor-selected sentences occurring under the heading word to the total number of sentences occurring under it. The frequency of occurrence of the heading words in the sample was also listed as a guide to the reliability of the selection ratio.

Step 4. Assignment of Weights. The data of Steps 2 and 3 were then examined and a weight was assigned to each heading word. It was found that the selection ratios confirmed the linguistic appraisal of word importance. Additions and deletions were made: the former by listing variants of important words or meaning equivalents, the latter by the criterion of low frequency. The final Heading Dictionary comprised 90 words (see Exhibit 5).

Ordinal Weights. Some 300 sentences were picked at random from the Heterogeneous Corpus and selection ratios computed. The data are as follows:

Ordinal Characteristics	Selection Ratio
First sentence of first paragraph	.60
Intermediate sentence of first paragraph	.35
Last sentence of first paragraph	.38
First sentence of intermediate paragraph	.44
Intermediate sentence of intermediate paragraph	.20
Last sentence of intermediate paragraph	.17
First sentence of last paragraph	.38
Intermediate sentence of last paragraph	.23
Last sentence of last paragraph	.24

These data served to establish weights for the ordinal characteristics: first paragraph, last paragraph, first sentence, last sentence. We may think of the weights of intermediate sentences of paragraphs as being zero. The adopted weights (see Exhibit 22) for the additive weight system reflect the gross behavior of the statistical data.

(3) Title Weights. The present research involved the creation of two new techniques relying upon two sentence characteristics not previously investigated. The first of these is founded upon words of the title and subtitles (i. e., headings) and hence is called the Title method. It is based upon purely linguistic considerations stemming from Principle 1 mentioned in Section 3. 3. 2. Here we rely upon the fact that a skilled author conceives of the title as circumscribing the subject matter of the document. Thus, whether or not the author incorporates his own abstract at the beginning of the document, he in effect conceives a title that can be viewed as an abstract of the abstract of the document. Similarly, when he partitions the body of the document into major sections he summarizes, by choosing the proper words to form his subtitles. Thus, it is believed that the content words of title and subtitles contain an important source of clues for an automatic abstracting system. In an examination of this hypothesis, title and subtitle words were verified as being statistically relevant characteristics. The hypothesis that such content words are irrelevant can be rejected at the 1% level of significance.

The Title method then creates, by a computer program, a Title Glossary consisting of the content words of the title and subtitles of the document. Words in the main body are then matched against the Title words and a match awards each sentence a positive weight. The weights assigned to the words of the Title Glossary are based on the consideration of their effect in the combined weighting scheme of the four methods. It was decided that content words of the title should outweigh content words of the headings. Therefore, the former were initially given the weight of 20 and the latter a weight of 10. This assignment of weights however, led to a difficulty in the ranking of all sentences of the document when the Title method was used. This was due to 20 being an exact multiple of 10, which caused many ties among sentences weights.

To avoid the occurrence of numerous ties, the following device was employed. Title words were assigned the weight of 11, and heading words the weight of 7 (see Exhibit 22). This arithmetic trick is based upon the fact that 11 and 7 are relatively prime and thus the probability of a tie is extremely small.

(4) Key Word Weights. The principle of the Key method was the first one proposed for the creation of automatic abstracts.

The previous study used the following definition of Key words: Candidates for Key words were first selected as all non-Cue words that occurred in the top 25 per cent Cue-weighted sentences. The Key word candidates were then frequency-counted over this collection of sentences and ranked in order of frequency. Key words were defined as those candidates which totaled the first 100 occurrences and the Key weights were taken to be the frequency of occurrence over that collection. The present study led us to change both of these conditions so that in the present system Key words are chosen from among the top 10 per cent of the total number of words in the document, and secondly their frequency of occurrence is computed over all words of the entire text. The Key weight of a word is taken to be its frequency of occurrence in the document (see Exhibit 22).

It is felt that the change from constant threshold of 100 occurrences to a fractional threshold of 10 per cent is an improvement. Moreover, both statistical and linguistic investigation have supported the shift from the more narrow environment of high Cue-weighted sentences to the wider environment of all text words.

3.4 EXPERIMENTS

In this section we describe 17 experiments conducted in the course of the research. These experiments can be classified into 4 groups which represent project milestones. A chart of the experiments is given in Exhibit 23. The final production runs are

Experiment Group & No.	No. of Documents Abstracted	Output		Components				Cue Dictionary Version	Cue Weights			Definition of Key Word	Key Weight	Title[1] Weights		Ordinal Weights			
		Key List	Vertical Listing	C	K	T	L		Null	Bonus	Stigma			Title	Heading	Paragraph First[2]	Paragraph Last[2]	Sentence First	Sentence Last
Preliminary	1	×	×	×	×			Linguistic	0	1	-	Definition 1[3]	g_i[4]						
I-1	8	×	×	×	×			1	0	10	-10[5]	Definition 1	g_i						
I-2	8		×			×		1						10	5				
I-3	18	×	×	×	×	×		1	0	10	-10[5]	Definition 1	g_i	10	5				
II-4	18	×	×	×	×	×		3	0	10	-10	Definition 2[6]	f_i[7]	10	5				
II-5	18	×			×	×		3				Definition 2	f_i	10	5				
II-6	18			×		×		3	0	10	-10			10	5				
II-7	18	×		×	×			3	0	10	-10	Definition 2	f_i						
II-8	18			×				3	0	10	-10								
II-9	18	×			×			3				Definition 2	f_i						
II-10	18					×		3						10	5				
III-11[8]	20	×	×	×	×	×	×	3	0	10	-10	Definition 2	f_i	11	7	18	2	18	2
IV-12	20	×	×	×	×	×	×	4	0	10	-10	Definition 3[9]	f_i	11	7	18	9	18	9
IV-13	20			×				4	0	10	-10								
IV-14	20	×			×			4				Definition 3	f_i						
IV-15	20						×									18	9	18	9
IV-16	20			×		×	×	4	0	10	-10			11	7	18	9	18	9
IV-17	20					×		4						11	7				
Production Runs	211			×		×	×	4	0	10	-10			11	7				

NOTES:

1. Title word = all non-Cue title and heading words.
2. "First" and "Last" of main sections of the document.
3. Definition 1. Key word = member of the shortest list of non-Cue words that comprises 100 word occurrences in the 25 percent of the sentences with the highest Cue weights.
4. g_i = frequency of Key word in the 25 percent of the document sentences with the highest Cue weights.
5. System did not have the capability at this time to treat negative Cue weights, but Stigma words were carried in the dictionary.
6. Definition 2. Key word = member of the shortest list of non-Cue words that comprise 100 word occurrences in the document.
7. f_i = frequency of Key word in document.
8. New format (selected sentences meshed with document headings) first tested.
9. Definition 3. Key word = member of the shortest list of non-Cue words that comprises 10 percent of the number of word occurrences in the document.

Exhibit 23. Chart of Experiments

also represented as well as the initial system checkout. Thus the chart gives a history of the experiments, introduction of the various methods, and modification of the weighting rules.

A resume of the experiments now follows.

3.4.1 Preliminary Test

To facilitate the research planned in this study the computer programs used in the previous work were converted to the Space Technology Laboratories computer system (see Ref. 9) and the research output (e.g. vertical listing, etc.) was incorporated in the program. Upon completion of these tasks an initial system test was conducted. An exotic fuel document was taken through the system from copy, pre-edited version, keypunched deck, edit (transfer to magnetic tape), to machine abstract with satisfactory results. Format changes were made in the vertical listing printout (see Exhibit 16).

3.4.2 Experimental Cycles

Group I. The purposes of these experiments were:

(1) to study the effect of the Cue Dictionary, Version 1, on exotic fuel documents.

(2) to introduce the Title method both in isolation and combination with the Cue and Key methods.

(3) to generate a sample of automatic abstracts of exotic fuel documents for use in a pilot evaluation study (see Section 3.5). The results of the Cue Dictionary study are reported in Section 3.3.4. A review of the Title method output confirmed the conjecture that title and heading words play a significant role in the abstracting. A review of the Key word lists led to a modification of the Key word definition.

Group II. After the third Cue Dictionary compilation (see Section 3.3.4) and modification of the Key word definition, it was decided to do a more complete series of experiments. The purpose was

(1) to study the effect of the Cue Dictionary, Version 3.

(2) to study all possible combinations of the three methods, Cue (C), Key (K), Title (T):

   a. All three in combination (Experiment 4)

   b. In pairs (Experiments 5-7)

   c. In isolation (Experiments 8-10)

(3) to study the effect of the modified Key word definition.

A comparison of the machine abstracts with the target abstracts in Experiment 4 (C-K-T) revealed that the machine technique was selecting 37 per cent of the target abstract sentences. A study of the printouts led to the following decisions:

(1) The format of the abstracts should be changed to include all headings together with the selected sentences. Because of the heading structure of the technical reports under consideration, it was noted that: (a) if a machine abstract had a comparable structure it could more easily serve the screening function of abstracts, (b) this structure would help clarify the meaning of a sentence lifted out of context.

(2) The Title weights should be changed to give better discrimination between sentences. The arithmetic trick described in Section 3.3.4 was used.

(3) The Location method should be tested in combination with the Cue, Key, and Title methods before final resolution of the Key word definition (which was still unsatisfactory).

Group III. This consisted of a single experiment (11) for the purpose of studying the new format, using a corrected Version 3 of the Cue Dictionary,* and testing the new Location method in

*It was found that 51 Cue words had not been transferred to the Dictionary tape.

combination. A comparison with target abstracts showed that 44 per cent of the target abstract sentences were now being selected by the machine.* It was apparent that the Location factors made a significant improvement in the technique.

A review of the data Groups I, II, III suggested that another group of experiments would be adequate to choose the final abstracting techniques in accordance with the scope of the present effort. It was decided that

(1) The abstract format was satisfactory.

(2) The Cue Dictionary, Version 4 (obtained by the statistical revision of Version 3) was adequate.

(3) A satisfactory Key word list would be obtained by taking the shortest list of Key words that comprises 10 per cent of the total number of word occurrences in the document.

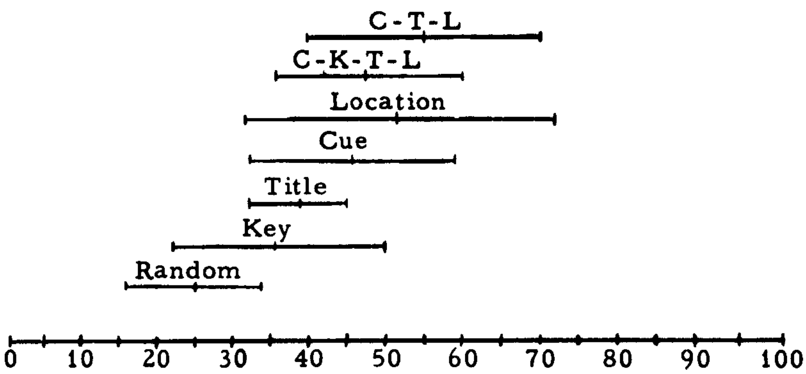
(4) The Title weights were satisfactory.

(5) The Heading Dictionary was satisfactory.

Group IV. These experiments were designed to test the above decisions. The results are given in two studies.

First, the percent of sentences in common with the target abstracts was computed. The mean percentages are shown below together with the intervals encompassing the mean plus or minus one standard deviation. The data for a random selection of 25 per cent of the sentences is given for comparison. It is to be noted that the C-T-L method has the highest mean value while the Key method in isolation has the lowest.

*Headings are not counted in the computation of "overlap" data.



On the basis of these data it was decided to omit the Key word component in the final abstracting system. The data confirms a hypothesis set forth previously* that Key words, while important for indexing, may not be so important for abstracting. This decision has important consequences for an abstracting system: considerable simplification can be achieved in the computer program if frequency-counting the entire text can be avoided.

The second study involved a detailed comparison of the abstracts produced by the C-T-L method with the target abstracts. We can uniquely classify every sentence in a document by considering all combinations of the properties: worthy to be in an abstract, in the target abstract, in the machine abstract, and the negations of these properties. Eight classes result, of which two define the most significant machine errors, namely:

> Type 1 Error: Sentence is worthy and in the target abstract, but is not in machine abstract.
>
> Type 2 Error: Sentence is not worthy and is not in the target abstract but is in the machine abstract.

Exhibit 24 gives the tabular breakdown of the analysis of the C-T-L abstracts of Experiment 16.

---

*Ref. 7

Document	Type 1 Errors: Sentences of Target Abstract Unmatched with Equivalent Machine-Selected Sentences	Sentences Selected by Machine					
		Total Sentences in Abstract	Agreements with Target Sentences	Additional Sentences Co-intensional with Target Sentences	Additional Worthy Sentences	Total Abstract-Worthy Sentences	Type 2 Errors: Sentences Which Are Not Abstract-Worthy
0003	5	12	5	2	3	10	2
0005	7	27	19	1	5	25	2
0006	5	15	10	4	0	14	1
0007	3	12	9	1	0	10	2
0008	8	12	3	2	2	7	5
0010	11	26	15	2	4	21	5
0011	3	11	9	2	0	11	0
0012	7	19	10	7	0	17	2
0014	2	5	2	0	3	5	0
0016	1	20	13	2	4	19	1
0018	9	25	15	0	7	22	3
0019	7	16	7	5	2	14	2
0023	0	6	5	1	0	6	0
0024	7	17	9	0	4	13	4
0042A	11	24	13	0	3	16	8
0045	8	21	13	0	5	18	3
0047	6	11	5	0	4		2
0052	5	14	9	0	4	13	1
0055	5	9	4	2	2	8	1
0056	6	9	2	1	0	3	6
Total:	116	311	177	32	52	261	50
Percent:	37%	100%	57%	10%	17%	$\boxed{84\%}$	16%

Exhibit 24. Analysis of C-T-L Selected Sentences

3.5 EVALUATION

A sample of 40 documents of the Test Library (documents not used in the abstracting experiments) was selected for evaluation. The purpose was to arrive at a gross evaluation of the quality of the machine product. To this end an evaluation procedure was designed based on (1) the definition of target abstract, and (2) the rating of the degree of similarity between two abstracts. Random extracts were used as controls. The Instructions to Raters are shown in Exhibit 25 and a Rating Form is shown in Exhibit 26.

3.5.1 Rating Procedure

(1) Uniform Presentation of Materials. The machine abstract, target abstract, and randomly generated extract were typed in the same format. Typing was necessary because machine printouts of the target abstract and random extract were not available. Headings were included as well as paragraph and sentence designations. The target abstract was identified, but the machine abstract and random extract were code designated.

(2) Similarity Rating. The raters judged the similarity between the target abstract and each of the other two according to the Instructions for Raters.

(3) Scoring. The ratings were scored by giving 4 points for complete similarity, 3 points for considerable similarity, . . . , and 0 points for no similarity.

A maximum possible score was next computed by considering the number of not-applicable information types (none gives a maximum of 24, 1 gives a maximum of 20, etc.). By dividing the total score by the maximum possible a normalized score was then obtained. This was interpreted as the degree of similarity between the abstracts and was recorded as a percentage.

(4) Participants. Two raters were used. The first rater was familiar with the 40 documents. The second was not associated with the project except for the evaluation test. The resulting ratings, reported below, justified the decision that a third rater was not required. The high consistency of the raters is shown by the small variation of the mean scores in three samples of 10 documents each.

	Target vs. Machine		Target vs. Random	
Samples	Rater 1	Rater 2	Rater 1	Rater 2
Sample 1 (n = 10)	72%	85%	39%	38%
Sample 2 (n = 10)	53%	56%	37%	30%
Sample 3 (n = 10)	66%	61%	32%	29%

The scores of the two raters were averaged for a single similarity score. Thus for each document two calculations were made: the percent similarity between Target and Machine, and percent similarity between Target and Random.

(5) Sampling and Assembling of Data. The documents were divided into 4 samples of 10 each. The purpose was to obtain a reliable mean similarity rating by working through these samples. It was found that 3 samples were sufficient. For each sample the mean similarity rating was computed as well as the standard deviation of the mean. The following table shows the standard deviations of the similarity ratings for cumulative samples.

INSTRUCTIONS FOR RATERS

You are asked to read a pair of abstracts of a document and then to judge the degree of similarity of the information contained in the two abstracts. It is not necessary that you understand the subject under discussion: only that you recognize the kind of information given.

Please compare the two abstracts with regard to the following six information types:

Subject Matter. Information indicating the general subject area that is the author's principal concern; i. e., what?

Purpose. Information indicating whether the author's principal intent is to offer original research findings, to survey or to evaluate the work performed by others, to present a speculative or theoretical discussion, or to serve some other main purpose; i. e., why?

Methods. Information indicating the methods used in conducting the research. Depending on the type of research, such statements may refer to experimental procedures, mathematical techniques, or other methods of scientific investigation; i. e., how?

Conclusions or Findings. Information indicating the results obtained in the research or the findings of the author.

Generalizations or Implications. Information indicating the significance of the research and its bearing on broader technical problems or theory.

Recommendations or Suggestions. Information indicating recommended courses of action or suggested areas of future work.

Using the following rating scale, place a check in the box on the rating form (see Exhibit 26) corresponding to one of the degrees listed below that best indicates the degree of similarity between the two abstracts·

Not applicable
No similarity
Moderate similarity
Considerable similarity
Complete similarity

Exhibit 25. Instructions for Raters

RATER:

Abstract No. ______ DOCUMENT: ______
vs. Abstract No. ______

SIMILARITY RATING FORM	Not Applicable	No Similarity	Slight Similarity	Moderate Similarity	Considerable Similarity	Complete Similarity
1. Subject Matter, i. e., what?						
2. Purpose, i. e., why?						
3. Methods, i. e., how?						
4. Conclusions or Findings						
5. Generalizations or Implications						
6. Recommendations or Suggestions						

REMARKS:

Exhibit 26. Rating Form

	Target vs. Machine	Target vs. Random
First Determination: Sample 1 (n = 10)	m = 79% s_m = 4%	m = 38% s_m = 4%
Second Determination: Samples 1, 2 (n = 20)	m = 67% s_m = 5%	m = 36% s_m = 4%
Third Determination: Samples 1, 2, 3 (n = 30)	m = 66% s_m = 3%	m = 34% s_m = 3%

Overlap data (i. e., percent agreement between Target and Machine) for cumulative samples 1, 2, and 3 gives 44 percent agreement between machine and target. Thus we have the following correspondence between similarity ratings and overlap data:

% Similarity	% Overlap	
100	100	(defined)
66	44	(computed)
34	25	(computed)

(6) Conclusions. (a) The mean similarity rating between the target and machine abstracts is 66 per cent; the standard deviation of the mean is 3 per cent. (b) The mean similarity rating between the target and the randomly generated abstracts is 34 per cent; the standard deviation of the mean is 3 per cent.

3. 5. 2 Comments on Evaluation Procedures

Because no attempt was made to evaluate the utility of the target abstracts, the above findings (that the machine abstract mirrors the target abstract at a 66% level) do not evaluate the utility of the machine abstracts. No conclusions about the utility of the machine abstracts can be drawn from the relative evaluation criteria utilized in this effort. However, it is interesting to analyze the reasons for the apparent useful qualities of the machine abstracts. A sentence-by-sentence analysis is summarized in Exhibit 24. Such an analysis is valuable as a supplement to the similarity rating because a high similarity may exist in specific sentences without coherence and a low similarity rating may be based on a poor representative of an information type.

The number of agreements with target sentences shows only part of the actual correspondence between the machine and human abstracts. When the number is expanded to include the machine-selected sentences cointensional with target sentences, the total represents the degree to which the machine process included the information of the target abstract. The category of "Additional Worthy Sentences" includes those which might be included in an abstract of unrestricted length. This group and the preceding two comprise the "Total of Abstract-Worthy Sentences", which may be taken to represent that part of the machine abstract which conforms to the content aspect of the working definition of "abstract". The average of 84 percent, even though uncorrected for redundancy in the machine abstract, seems to represent a highly promising achievement in the automation of the abstracting process.

The sentences selected by the machine process but not abstract-worthy (Type 2 errors) are extraneous detail and represent "noise". They clutter the abstract and often interfere seriously with coherence. To minimize this group should be one goal of future research. The sentences resulting in Type 1 errors represent information included in the target abstract but not in the machine abstract. Their significance can only be discovered by looking at the sentence in question.

Future research should involve statistical analysis of the two error-types with the purpose of modifying the program to minimize them. Study should be made to discover machine-recognizable clues to determine the proper length of an abstract. The extent to which redundancy appears in the machine abstract and ways of mechanizing its suppression should be investigated. Linguistic clues to coherence should be investigated and expressed in machine-recognizable form, perhaps in the form of a word-and-phrase dictionary indicating the need for selecting an antecedent sentence.

In short, the main differences between human and machine abstracts (see Section 3. 3. 1) can now be described and tabulated and procedures outlined to minimize them. However, in the last analysis, progress in the creation of automatic abstracts must be verified by the users.

REFERENCES

1. Proposal for a Study of Automatic Abstracting, Attachment III - Technical Discussion, P705-9U3, 2 December 1959, Ramo-Wooldridge, a Division of Thompson Ramo Wooldridge Inc., Canoga Park, California.

2. Study for Automatic Abstracting, Quarterly Memorandum, AF 30(602)-2223, 17 June 1960.

3. Study for Automatic Abstracting, Quarterly Memorandum, AF 30(602)-2223, 17 September 1960.

4. Proposal for Engineering Change "A" to Contract AF 30(602)-2223 (Study for Automatic Abstracting), Attachment II - Technical Discussion, P1214-0U2, 2 December 1960, Ramo-Wooldridge, a Division of Thompson Ramo Wooldridge Inc., Canoga Park, California.

5. Study for Automatic Abstracting, Quarterly Memorandum and Interim Technical Report, AF 30(602)-2223, C107-1U6, 17 March 1960, Ramo-Wooldridge, a Division of Thompson Ramo Wooldridge Inc., Canoga Park, California.

6. Study for Automatic Abstracting, Quarterly Memorandum AF 30(602)-2223, 17 June 1961.

7. Study for Automatic Abstracting, Final Report, AF 30(602)-2223 C107-1U12, 1 September 1961, Ramo-Wooldridge, a Division of Thompson Ramo Wooldridge Inc., Canoga Park, California.

8. Proposal for Engineering Change "B" to Contract AF 30(602)-2223 (Study for Automatic Abstracting), Attachment II - Technical Discussion, P1397-1U2, 1 October 1961, Ramo-Wooldridge, a Division of Thompson Ramo Wooldridge Inc., Canoga Park, California.

9. Study for Automatic Abstracting, Quarterly Memorandum, AF 30(602)-2223, 30 April 1962.

10. Study for Automatic Abstracting, Quarterly Memorandum, AF 30(602)-2223, 30 July 1962.

11. Study for Automatic Abstracting, Quarterly Memorandum, AF 30(602)-2223, 30 October 1962.

12. H. P. Edmundson and R. E. Wyllys, Automatic Abstracting and Indexing--Survey and Recommendations, Communications of the Association for Computing Machinery, vol. 4, no. 5, pp. 226-234, 1961.

13. H. P. Edmundson, "A Statistician's View of Linguistic Models and Language Data Processing," Natural Language and the Computer, P. Garvin (editor), to be published by McGraw-Hill.

14. H. P. Edmundson, "Problems in Automatic Abstracting," Joint Man-Computer Indexing and Abstracting, 20 November 1962, Session 13, First Congress on the Information System Sciences.

15. J. L. Kuhns, "An Application of Logical Probability to Problems in Automatic Abstracting and Information Retrieval," Joint Man-Computer Indexing and Abstracting, 20 November 1962, Session 13, First Congress on the Information System Sciences.